

International Classification of Diseases (ICD)[†]

Yi Hong* and Marcia Lei Zeng**

*MedeAnalytics, Inc., Department of Professional Services, Dallas, Texas, USA,

<ylphhong@yahoo.com>

**School of Library and Information Science SLIS, Kent State University, Kent, Ohio 44242-0001, USA,

<mzeng@kent.edu>



Yi Hong holds a doctorate degree in biomedical and health informatics and a master's degree in information studies from University of Wisconsin-Milwaukee. She is a data architect at the MedeAnalytics, a health data analytics company in USA. She served as clinical informatics analyst at Harris Corporation and worked at Medical College of Wisconsin as a clinical systems specialist before moving to her current position. Her research interests focus on health knowledge organization systems, biomedical terminology services, information architecture, health data modeling and analysis.



Marcia Lei Zeng is professor of Information Science at Kent State University. She holds a Ph.D. from the School of Computing and Information at University of Pittsburgh (USA) and M.A. from Wuhan University (China). Her major research interests include knowledge organization systems (taxonomy, thesaurus, ontology, etc.), linked data, metadata and markup languages, smart data and big data, database quality control, semantic technologies, and digital humanities. Her research projects have received funding from the National Science Foundation (NSF), Institute of Museum and Library Services (IMLS), OCLC Online Computer Library Center, Fulbright, and other foundations. Dr. Zeng has chaired or served on committees, working groups, and executive boards for IFLA, SLA, ASIS&T, NISO, ISO, DCMI, ISKO, and W3C. She is serving as a DCMI Governing Board and Executive Board Member of ISKO.

Hong, Yi and Marcia Lei Zeng. 2022. "International Classification of Diseases (ICD)." *Knowledge Organization* 49(7): 496-528. 120 references. DOI:10.5771/0943-7444-2022-7-496.

Abstract: This article presents the history, contents, structures, functions, and applications of the *International Classification of Diseases* (ICD), which is a global standard maintained by the World Health Organization (WHO). The article aims to present ICD from the knowledge organization perspective and focuses on the current versions, ICD-10 and ICD-11. It also introduces the relationship between ICD and other health knowledge organization systems (KOSs), plus efforts in research and development reported in health informatics. The article concludes that the high-level effort of promoting a unified classification system such as ICD is critical in providing a common language for systematic recording, reporting, analysis, interpretation, and comparison of mortality and morbidity data. It greatly enhances the constancy of coding across languages, cultures, and healthcare systems around the world.

Received 9 October 2022; Accepted 10 October 2022

Keywords: controlled vocabularies, health sciences, medical KOS

[†] Derived from the article titled "International Classification of Diseases (ICD)" in the ISKO Encyclopedia of Knowledge Organization, Version 1.0 published 2022-08-29. Article category: KOS, specific (domain specific).

1.0 Introduction

The *International Classification of Diseases* (ICD) is a global standard maintained by the World Health Organization (WHO) for classifying diseases and reporting health conditions, as well as a foundation for health trends and statistical analysis. It is steeped in history, international cooperation, and improvement over the last 150 years (Salcido 2015). As a health knowledge organization system (KOS)

that defines diseases, disorders, injuries, and other related health conditions in a classificatory structure, ICD has been gradually developed as a diagnostic classification standard for epidemiology, health management, clinical, and research purposes. Its development has augmented efforts to enhance its coverage and implement a unified notation system for usage across languages, cultures, and global healthcare systems. "The purpose of the ICD is to allow the systematic recording, analysis, interpretation and comparison of mor-

tality and morbidity data collected in different countries or regions and at different times” (World Health Organization 2022a, Sec.1.1).

Understanding ICD is vital to working with health data since it is used worldwide for classifying diseases and other health problems and encoding diagnoses of diseases in an alphanumeric coding system for all general epidemiological and many health-management purposes. ICD has played a critical role in the exchange of health information, especially during a pandemic (Tsui et al. 2002; Moore et al. 2011; Gerbier-Colomban 2014; Ben-Ezra et al. 2021). As pointed out by the current *ICD-11 Reference Guide* (World Health Organization 2022a, Sec. 1.1), while most widespread use of ICD over time and around the world is for mortality statistics, it is also used for classification of clinical documentation to provide standardized, language independent information for morbidity use, such as resource allocation, case-mix, patient safety and quality of care alongside primary care and research. Moreover, ICD and its descriptions are used as a framework in legislation. As the most widely used classification in healthcare systems around the world, ICD has been followed as the standard by other health KOSs, with additional extension (both in depth and width), incorporation (based-on or derived from), crosswalking, and complementary modeling efforts.

This article presents the history, contents, structures, functions, and applications of the ICD, focusing on the current versions, ICD-10 and ICD-11. The article aims primarily to present ICD from the knowledge organization (KO) point of view, facilitating the understanding of IDC foremost, which will then lay the foundation to further apply, analyze, evaluate, and even assist in its development. This article also introduces the relationship between ICD and other health KOSs, involving those considered as the WHO Family of International Classifications (WHO-FIC) and the ones developed and used in different processes and countries. Efforts in research and development reported across health informatics are given at an introductory level. Figures have been created to help with the interpretation of major viewpoints and approaches. Accompanying examples have been used, with sources provided.

With the evolution of ICD from its early versions entitled *International List of Causes of Death* (ICD-1 to ICD-5) to *International Statistical Classification of Diseases, Injuries, and Cause of Death* (ICD-6) and onward, the full titles of ICD revisions have been modified while still including International Classification of Diseases in their titles, as shown in Figure 2. Regarding these given various full titles in the printed and online versions, this article follows the WHO’s ICD website and uses “ICD” along with “Revision #” as the acronym for this classification (Moriyama et al. 2011; World Health Organization 2021a). In other words, “ICD-#” will be consistently used as a pattern regardless of

the variance of full titles, parallel to the style used in publications, such as the one by DiChiara (2015), titled “ICD-1 to ICD-11 Timeline Highlights Healthcare’s Evolution”.

2.0 Early Development of disease classifications

The concept of ICD is rooted in the theory of nosology, the systematic classification of diseases. “Nosologic classification began in antiquity, resulting out of the need for nurses, physicians, epidemiologists, and public health entities to classify and make sense of cause of death and morbidity; the parallel use of these data can be traced to the 15th century in Italy, as a result of the ‘great pandemics of plague’” (Salcido 2015). The idea that diseases are directly identifiable through their symptomatology was found in the revolutionary work of Paracelsus (1493-1541). Paracelsus’ suggestion that natural symptoms may provide direct evidence leading to a probable diagnosis or classification of a disease, despite a potentially unobservable cause of the illness, was evident in practice by the 17th century when nosology took form. Throughout the 18th century, this pathological view was guided by the concept that a specific cause could be identified for all episodes of ill health, and physicians concentrated on the observation of symptoms and the categorization of diseases (Encyclopedia.com. 2019).

Nosologia Methodica, a comprehensive treatise by François Boissier de la Croix de Sauvages (1706-1767), was credited as the first attempt to classify diseases systematically by the eminent Australian statistician George Knibbs (1929). However, the *Synopsis Nosologiae Methodicae*, published in 1785 by William Cullen (1710-1790) of Edinburgh, became more popular in general use in the public services at the beginning of the 19th century due to the simplicity and merits of its classification (World Health Organization [2004], Moriyama et al. 2011).

William Farr (1807-1883) conducted constant studies on existing nomenclature and classifications and emphasized the importance of a uniform statistical classification of causes of death. He was the first medical statistician to make the best possible use of the imperfect classifications of diseases available at the time. In 1839, his effort at a statistical classification appeared in the *First Annual Report of the Registrar General of Births, Deaths, and Marriages in England* (Eyler 1979).

In 1853, William Farr and Marc d’Espine (1806-1860), who created a statistical nosology in Geneva, were requested to draft an internationally applicable, uniform classification of causes of death at the first International Statistical Congress. Farr and d’Espine presented two different lists following the divergent principles at the next congress in 1855. Farr classified diseases under five groups (epidemic diseases, constitutional [general] diseases, local diseases arranged according to anatomical site, developmental diseases, and dis-

eases directly resulting from violence) while d'Espine classified diseases according to their nature (gouty, herpetic, haematic, etc.) (Moriyama et al. 2011; Lewes 1988). The congress accepted a compromise list of 139 rubrics. The 1860 International Statistical Congress meeting in Paris discussed hospital statistics and adopted a complete statistical layout for classifying hospital cases, using a list of causes said to be based on the 1855 Paris list and the same one used by Farr at the General Register Office for England and Wales. Florence Nightingale (1820-1910), who attempted to classify morbidity and mortality using evidence-based technique during the Crimean War (1853-1856) period, was the driving force of this discussion.¹ She proposed a very elaborate plan aimed at statistically demonstrating how improved sanitary conditions and better schooling reduced mortality, illness, and even criminal behavior (Moriyama et al. 2011; Lilienfeld 2007; McDonald 2001; Salcido 2015). Acknowledged in the *History of the Development of the ICD* (World Health Organization [2004]), the list was subsequently revised in accordance with Farr's model in 1864, 1874, 1880, and 1886, therefore prevailing as the basis of the *International List of Causes of Death* with the principle of classifying diseases by anatomical site.²

In 1869, the *Nomenclature of Diseases* by the Royal College of Physicians of London was published with a listing of approved names of diseases in English, Latin, German, French, and Italian, including synonyms for English names and definitions. It was the first authoritative source-of-disease KOS that was a nosological grouping rather than a classification (Robb-Smith 1969). According to Thompson and Hayden (1961), the *Nomenclature of Diseases* was re-

vised in 1885, 1896, 1906, 1918, 1931, 1947, and 1959 to provide an authoritative source of medical terminology for British physicians and was followed shortly by the American Medical Association (AMA)'s nomenclature of disease.

In 1891, a committee chaired by Jacques Bertillon (1851-1922) was assigned to create a classification of causes of death by the International Statistical Institute (ISI). Bertillon introduced an alphanumeric method of disease classification, which incorporated disease by anatomical site and cause of death (Salcido 2015). In 1893, the report of the committee chaired by Bertillon was submitted and adopted by ISI at its meeting in Chicago. Publication of this report was the origin of the *International List of Causes of Death* (Moriyama et al. 2011). The classification proposed by Bertillon's committee was based on the classification of causes of death used by Paris, which adopted Farr's principle of distinguishing between general diseases and those localized to an organ or anatomical site. The main headings of Bertillon's list are shown in Figure 1.

The Bertillon classification of causes of death had been embraced across North America (United States, Canada, Mexico), several parts of South America, and in some cities in Europe by the time of ISI's 1899 meeting (Moriyama et al. 2011).

3.0 A brief history of ICD

Considering the *International List of Causes of Death* which the International Statistical Institute adopted in 1893 as the first international classification edition, ICD has since been revised and published in a series of revisions to reflect timely

- | | |
|--------------|--|
| I. | General diseases |
| II. | Diseases of nervous system and sense organs |
| III. | Diseases of circulatory system |
| IV. | Diseases of respiratory system |
| V. | Diseases of digestive system |
| VI. | Diseases of genitourinary system |
| VII. | Puerperal diseases |
| VIII. | Diseases of skin and annexes |
| IX. | Diseases of locomotor organs |
| X. | Malformations |
| XI. | Diseases of early infancy |
| XII. | Diseases of old age |
| XIII. | Effects of external causes |
| XIV. | Ill-defined diseases |

Figure 1. Main headings of Bertillon's list (Moriyama et al. 2011, 12).

advances in health and medical science. The history of ICD is well carried by a summary, *History of the Development of the ICD*, provided by the World Health Organization [2004] and a book, *History of the Statistical Classification of Diseases and Causes of Death* (Moriyama et al. 2011), published by the National Center for Health Statistics of the U.S. Centers for Disease Control and Prevention (CDC), comprising the following momentous outcomes:

The *International List of Causes of Death, Revision 1* (ICD-1) was released at the first International Conference for the Revision of the *International List of Causes of Death* in Paris on August 18, 1900. It adopted the same structure as Bertillon's list, except that the first main heading, "General diseases", was replaced by two subheadings, one for "Epidemic Diseases" and the other for "Other General Diseases". The diseases were identified and sorted by Arabic numerals, which are integers, except for the entry of 174.1 for "Other specified diseases", as shown in the digitized document provided by Wolfbane Cybernetic (2001-) on the web, at <http://www.wolfbane.com/icd/icd1h.htm>.³ The usage of ICD-1 quickly spread after it was translated from French into several other languages. By 1909, ICD-1 was in use throughout the world (Moriyama et al. 2011).

The *International List of Causes of Death, Revision 2* (ICD-2) was released in 1909. "The translation in English of the Second Decennial Revision, published by the United States Department of Commerce and Labor in 1910, was entitled *International Classification of Causes of Sickness and Death*" (World Health Organization [2004], 4). The English version used the words "nomenclature" and "classification" to describe the disease list for the first time. In addition, the English version contained an expanded alphabetical index, which showed the source of the items and gave the rubric numbers of both the detailed and abridged lists (U.S. Bureau of the Census 1911). The revisions continued with ICD-3 to ICD-5 released through wider international collaborations (World Health Organization 1948).

For the sixth revision, the International Health Conference (IHC) held in New York City in June and July 1946 entrusted the Interim Commission of the WHO with the responsibility of reviewing the existing machinery and of undertaking such preparatory work as may be necessary in connection with: (i) the next decennial revision of *The International Lists of Causes of Death* and (ii) the establishment of *International Lists of Causes of Morbidity*. Its Expert Committee's resulting work, the sixth revision titled *International Classification of Diseases, Injuries, and Causes of Death* (World Health Organization [2004]), sought comments and suggestion and was circulated to national governments that were preparing morbidity and mortality statistics. The sixth revision conference endorsed publication of the *Manual of the International Classification of Diseases, Injuries and Causes of Death*. The International Classifica-

tion, including the Tabular List⁴ of Inclusions defining the content of the categories, was incorporated together with the form of the medical certificate of cause of death, the rules for classification, and the special lists for tabulation. In 1948, the First World Health Assembly adopted ICD-6, which incorporated morbidity for the first time (Moriyama et al. 2011, World Health Organization [2004]).

ICD-6 presents a milestone in the history of ICD regarding both content and range of application. As the first version supported morbidity reporting, it was also the first to have the term "classification" in the title (World Health Organization 1948). The scope of ICD-6 expanded to apply explicitly to morbidity as well as mortality. Mental disorders were added for the first time to reflect the need for coding mental conditions, while the concept of a primary cause of death for tabulation was refined and operationalized. The legal authority of the classification was strengthened and formalized. Besides the major changes in scope and sections, the notation system was also significantly revised. For the first time, ICD-6 employed three-digit Arabic numerals for class/subclasses and decimal points for subsections, which reflected the hierarchical structure of the classification (see the lower right of Figure 2 below and find components from the website of digitized ICD versions provided by Wolfbane Cybernetic 2001-).

Figure 2 presents a timeline of ICD's development and implementations. The date ranges listed in the figure are the dates that the various revisions of the ICD were released or adopted and came into effect, as itemized in the book by Moriyama et al. (2011) and the "ICD History" (World Health Organization 2021a) section on the official ICD website.

The milestone change from the term "list" to "classification" can be seen in ICD-6. Since then, all releases of ICD (ICD-7 to ICD-11) have been from the WHO, and ICD has been used worldwide as a diagnostic classification standard. The title of the ICD has been designated in both the full title as well as the commonly referred to title, *International Classification of Diseases*. In general, the development of the ICD can be seen in three major portions:

- ICD-1 to ICD-5, establishing a common "term list"
- ICD-6, the turning point, with the title of "classification" and the WHO's action
- ICD-7 to ICD-11, the WHO's International Classification of Diseases

ICD-6 started a new era in international cooperation for vital and health statistics. The WHO began to lead the ICD revision and coordinating the statistical activities among the national committees of various countries. These national committees served as a link between the national statistical institutions and the WHO: the committees investigated

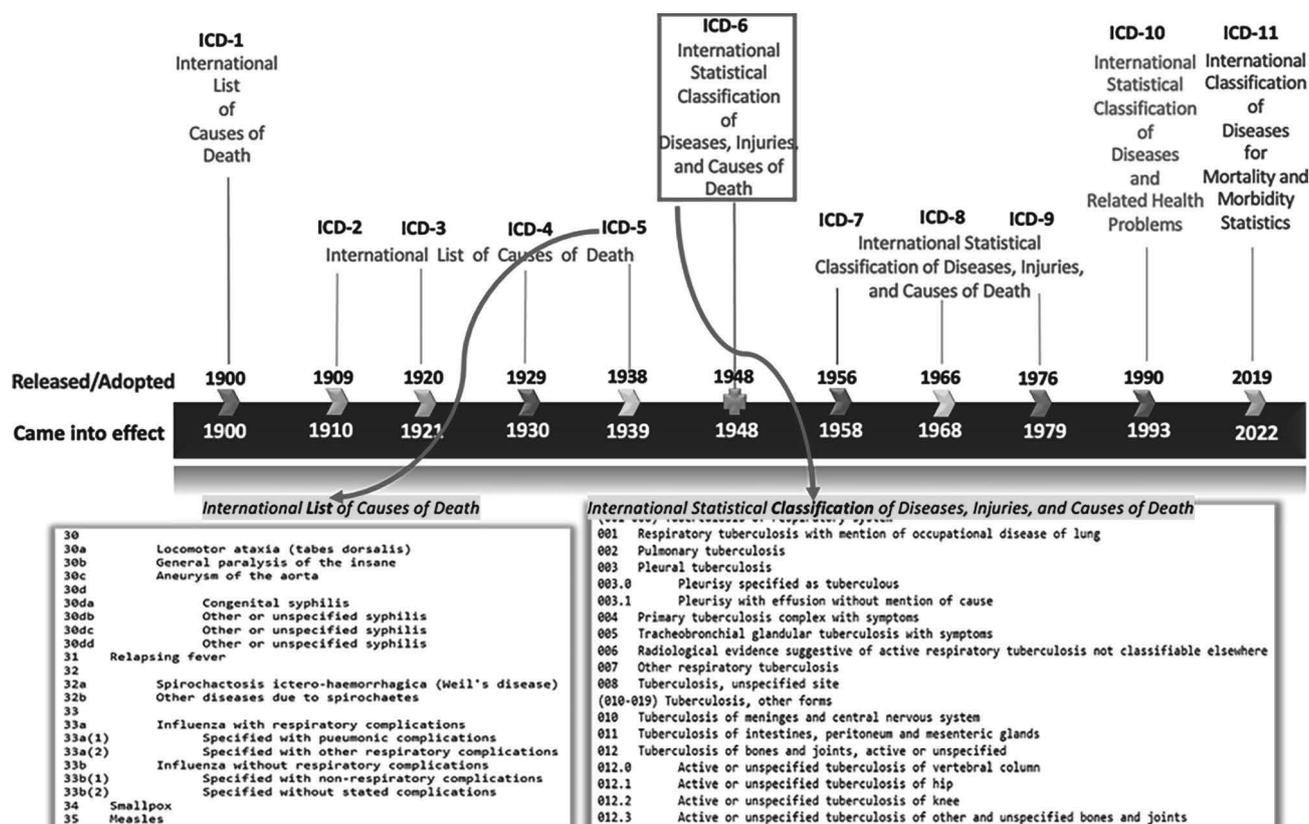


Figure 2. History of ICD revisions.

(Timeline created by the authors based on Moriyama et al. 2011 and World Health Organization 2021a, with the supplemental screenshots of partial ICD-5 and ICD-6 taken from Wolfbane Cybernetic 2001-)

statistical problems within public health records and made their research results available to the WHO. In 1951, the first WHO Center for Classification of Diseases was established at the General Register Office of England and Wales. The center served as a clearinghouse to solve problems in using ICD and assist the WHO Secretariat with ICD development in a setting where data were available for testing revision proposals (World Health Organization [2004]).

ICD-7, ICD-8, and ICD-9 (came into effect in 1958, 1968, and 1979) retained the basic structure of the ICD-6 and continually extending the subcategories or subdivisions. While the adaptations of ICD-7 influenced the development of ICD-8 to meet the needs of hospitals, the use of ICD for indexing hospital medical records increased rapidly. Furthermore, some countries furnished national adaptations that provided additional details for coding hospital and morbidity data during the years that ICD-7 and ICD-8 were in force. For example, the ICD was adapted for indexing of hospital records and operation classification and served as the basis for coding diagnostic data for both official morbidity and mortality statistics in the United States (U.S. Department of Health, Education and Welfare 1959;

1968). The ICD-9 came into effect in 1979. "For the first time, the ninth revision contained guidance on recording and coding for morbidity and specifically for the selection of a single condition for presentation of morbidity statistics" (World Health Organization 2016, Vol.1, 20).

When ICD-9 was released in 1978, the *International Classification of Procedures in Medicine* (ICPM) was also published by the WHO (World Health Organization 1978). Intended as a supplement to ICD-9, ICPM focused on procedures following the structure and system of coding used in ICD and covered diagnostic, laboratory, radiological, preventive, surgical, drug, and other therapeutic or ancillary procedures. The *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM), an official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States (U.S. National Center for Health Statistics 2021), was adapted by the U.S. National Center for Health Statistics (NCHS) and has been updated annually since 1996.

The 10th Revision of the ICD conference, held in 1989, recommended the proposed revised chapters, with their 3-character categories and 4-character subcategories, plus the

Short Tabulation Lists for Morbidity and Mortality constitute the ICD-10. The World Health Assembly adopted ICD-10 in May 1990 bringing it into force on January 1, 1993. The full title of ICD-10 is *International Statistical Classification of Diseases and Related Health Problems*. “While the title has been amended to make clearer the content and purpose and to reflect the progressive extension of the scope of the classification beyond diseases and injuries, the familiar abbreviation ‘ICD’ has been retained” (World Health Organization 2016 Vol.1, 1). ICD-10 was translated into the official languages of the United Nations as well as other languages. The official updates to the published volumes of ICD-10 have been available as annual lists of changes since 1996. The updates are approved annually at the October meeting by the heads of the WHO at Collaborating Centers for the Family of International Classifications (World Health Organization [1996-]). Three volumes comprise the ICD-10 Codebook (World Health Organization 2016):

- Volume 1, the Tabular List, contains the main classifications, including an alphanumeric listing of diseases, disease groups, and health related problems. It also contains inclusion and exclusion notes and some coding rules.
- Volume 2, the Instruction Manual, provides guidance and instructions regarding how to use the classification to code death certificates, hospital medical records, and other forms of health information.
- Volume 3, the Alphabetical Index, is an alphabetical list of the diseases and conditions which have codes in the Tabular List.

ICD-10, available in 43 languages and used in over 100 countries, also has more than two dozen modifications (Grove and Jakob 2018). Several national clinical modifications of ICD-10 have been developed, with permission from the WHO, to better suit the needs of morbidity reporting, usually by adding more details. A review by Jetté et al. (2010) contained the following: Australian Modification: ICD-10-AM (in use in Australia, New Zealand, Ireland, Slovenia), Canadian Enhancement (ICD-10-CA), France Clinical Addendum to ICD-10, German Modification (ICD-10-GM), Korean Modification (ICD-10-KM), Thai Modification, Thailand (ICD-10-TM), US Clinical Modification (ICD-10-CM), etc. Other notable modifications were developed for Estonia, the Netherlands, the Philippines, Singapore, Sweden, and Switzerland. All modifications to ICD-10 must conform to WHO conventions for ICD. The WHO applies certain restrictions to limit the types of modifications to ICD to prevent changes in the concepts and meanings of three-digit categories and four-digit codes (Terron Cuadrado 2019).

The 11th Revision, ICD-11, released in June 2018 (World Health Organization 2018) was officially endorsed by all WHO members during the 72nd World Health Assembly in May 2019 (World Health Organization 2019a). ICD-11 is a vast expansion on ICD-10, updated to reflect vital advances in life science and medicine for the 21st century, and enhanced to integrate with electronic health information systems. The definitive version of ICD-11 is available online, from <https://icd.who.int/browse11/l-m/en>, with the title *ICD-11 for Mortality and Morbidity Statistics* (ICD-11 MMS). ICD-11 is multipurpose where the structure is defined in a linearization, which incorporates properties and attributes focusing on mortality, morbidity, the degree of primary care, research, and public health (World Health Organization 2022a, Sec.1.1). Its new contents and tools provide approaches for reporting, coding, selecting, and tabulating conditions for different use cases, and include more clinical details to improve usability with multilingual support. The ICD-11 ensures consistency with traditional use cases of earlier ICD versions because it has been built with the past revisions in mind. Past data analyses based on older versions of ICD can be linked to analyses of data based on ICD-11.

4.0 Current versions of ICD (ICD-10 and ICD-11)

ICD has been published and maintained by the WHO since ICD-6 and is used worldwide for health statistics, disease classification, diagnostic coding, medical billing, reimbursement systems, and automated decision support in healthcare. It is one of the three primary reference classifications maintained by the WHO approved for international reporting and covers the main parameters of health (World Health Organization 2021a). The 10th Revision, with the full title *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), was adopted in May 1990, came into effect in 1993 and has been used in healthcare systems worldwide. The 11th Revision, entitled *ICD-11 for Mortality and Morbidity Statistics* (ICD-11 MMS), was adopted in May 2019 and officially started implementation on January 1, 2022, almost three decades after ICD-10's use. It is referred to as “ICD-11” (instead of “ICD-11 MMS”) by the WHO ICD homepage <https://icd.who.int>, the materials provided through its website, and its reference guide. Both ICD-10 and ICD-11 provide very rich reference guides (World Health Organization 2016, 2022a), which will be used in this section.

4.1 Content and chapter structure

4.1.1 Content

In the ICD-10's *Instruction Manual* (World Health Organization 2016 Vol.2) Sec. 2.3 on “General principles of dis-

ease classification” started with this 1856 statement by William Farr:

Classification is a method of generalization. Several classifications may, therefore, be used with advantage; and the physician, the pathologist, or the jurist, each from his own point of view, may legitimately classify the diseases and the causes of death in the way that he thinks best adapted to facilitate his inquiries, and to yield general results.

The Manual indicated that a statistical classification of diseases must be confined to a limited number of mutually exclusive categories that are able to encompass the whole range of morbid conditions.

This is consistently addressed within the *ICD-11 Reference Guide* section “1.2.1 Taxonomy” as well (World Health Organization 2022a, Sec.1.2.1). As a statistical classification of diseases, ICD follows the parameter that it must be confined to a limited number of mutually exclusive categories which are able to encompass the complete range of morbid conditions. The categories are chosen to facilitate the statistical study of disease phenomena. In determining whether an entity qualifies to become a unique category, the following measures are applied:

1. Epidemiological evidence: frequency analyses of coded mortality and morbidity data
2. Clinical evidence: disease evidence provided by the medical specialties
3. Granularity: minimum detail reported and useful in mortality or primary care
4. Continuity: preservation of the level of detail pre-existing in ICD
5. Parsimony: the need to limit the number of categories for international mandatory reporting.

The Content model of ICD-11 is a structured framework that defines each entity found in the ICD in a standard way. The purpose of the Content model is to present the background knowledge that provides the basis for the description of each ICD entity in a systematic way and allows for computerization (Sec. 3.4). Each ICD entity, which might become a grouping, a category, or just a term that is, for example, listed in the index, can be seen from different dimensions. The main properties of the Content model are: 1) ICD Concept Title; 2) Hierarchy, Type and Use; 3) Textual Definition(s), with a description and a definition; 4) Terms, including index terms, Synonyms Inclusion terms, and Exclusion Terms; and 5) Clinical Descriptions, which has the most comprehensive and inclusive perspectives.

For each entity, four levels of descriptor information are given in the ICD-11 Content model (Sec.3.4) in order to

enhance the comparability, consistency, and interpretation of coded information for everyone, everywhere: 1) Fully Specified Term; 2) Short Description; 3) Additional Information; and 4) Clinical or Diagnostic Criteria. Diagnostic Criteria contains the core diagnostic information necessary and sufficient to describe a category and enables the digital representation of the diagnostic algorithms using standardized terminology and other elements as appropriate.

Getting into the Definition of disease, a set of relevant aspects are drawn from the pattern below and are defined as “properties” (Sec.1.2.1.1). A disease is a set of dysfunctions in any body system defined by: 1) Symptomatology or manifestations; 2) Aetiology (an underlying explanatory mechanism); 3) Course and outcome; 4) Treatment response; 5) Linkage to genetic factors; and 6) Linkage to environmental factors.

The categories in ICD are mutually exclusive, jointly exhaustive, and linked to a monohierarchical tree. When an entity may be correctly classified in two different places, e.g., by site or by aetiology, indications of multiple parenting are provided through notes (‘Excludes’ or ‘Code elsewhere’) or specially designed displays. Every time an entity is parented elsewhere, it will continue to show the code from the primary parent. All ICD-11 categories include separate information on anatomy, etiology, and other aspects that can be accessed for browsing and searching purposes online through the ICD-11 browser <https://icd.who.int/browse11/>. The factsheet (World Health Organization 2021-) indicates that the full online versions in other standard languages of the WHO are available in Arabic, French, Mandarin, and Spanish, while Russian and 20 more languages are underway.

The hierarchical structure of the ICD allows considerable flexibility for other possible tabulations. The special tabulation lists are designed for the aggregation and reporting of coded data, to allow national and international comparisons between different countries or regions (World Health Organization 2022a, Sec.2.25). Both ICD-10 and ICD-11 have recommended special tabulation lists. Additional special tabulations can be derived from the new multiple parenting technique, e.g., all WHO notifiable diseases, listing all conditions that are assigned to the relevant section of the infectious diseases chapter (Sec.3.1).

4.1.2 Chapter structure

ICD is a variable-axis classification. The term “chapter” is used for the main classes of the classification. The distinction between the “special groups” chapters and the “body systems” chapters has practical implications for understanding the structure of the classification, for coding to it, and for interpreting statistics based on it (World Health Organization 2022a, Sec.1.2.2).

In general, the structure was developed from early international discussion on classification structure and the structure that was proposed by William Farr: epidemic diseases, constitutional or general diseases, local diseases arranged by site, developmental diseases, and injuries. The structure has stood the test of time and, though in some ways arbitrary, is still regarded as more useful for general epidemiological purposes than any of the alternatives tested (Sec.1.2.2). The conservation of the structure acknowledges the need for stability while allowing the incorporation of additional sections. These “body systems” chapters are maintained in the current ICD chapters. Meanwhile, some “special groups” chapters are also included to bring together conditions that would be inconveniently arranged for epidemiological study if they were to be scattered, such as in a classification arranged primarily by anatomical site. The first three chapters of both ICD-10 and ICD-11, shown in Figure 3, and some other chapters are considered “special groups” chapters.

ICD-10 has 22 chapters. To facilitate the statistical study of disease phenomena, every disease or morbid condition must have a well-defined place in the list of categories. Consequently, there are residual categories throughout the clas-

sification for other and miscellaneous conditions that cannot be allocated to the more specific categories (World Health Organization 2016 Vol.2, 14).

Figure 3 shows the current vision of ICD-10 available online. Many chapters in ICD-10 are associated with particular body systems. For example, Chapter X is for diseases of the respiratory system, while Chapter XI relates to diseases of the digestive system. Parallel to these “body systems” chapters, the “special groups” bring together conditions that would be inconveniently arranged for epidemiological study were they to be scattered, such as in a classification arranged primarily by anatomical site (World Health Organization 2016 Vol.2, 13-14). The chapters from I-XVII cover diseases and other morbid conditions. Chapter XVIII covers symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified. Chapter XIX relates to injury, poisoning, and certain other consequences of external causes. Chapter XX permits the classification of environmental events and circumstances as the cause of injury, poisoning and other adverse effects. Chapter XXI “Factors influencing health status and contact with health services,” is for the classification of data explaining the reason for contact with health-care services for a person not currently sick,

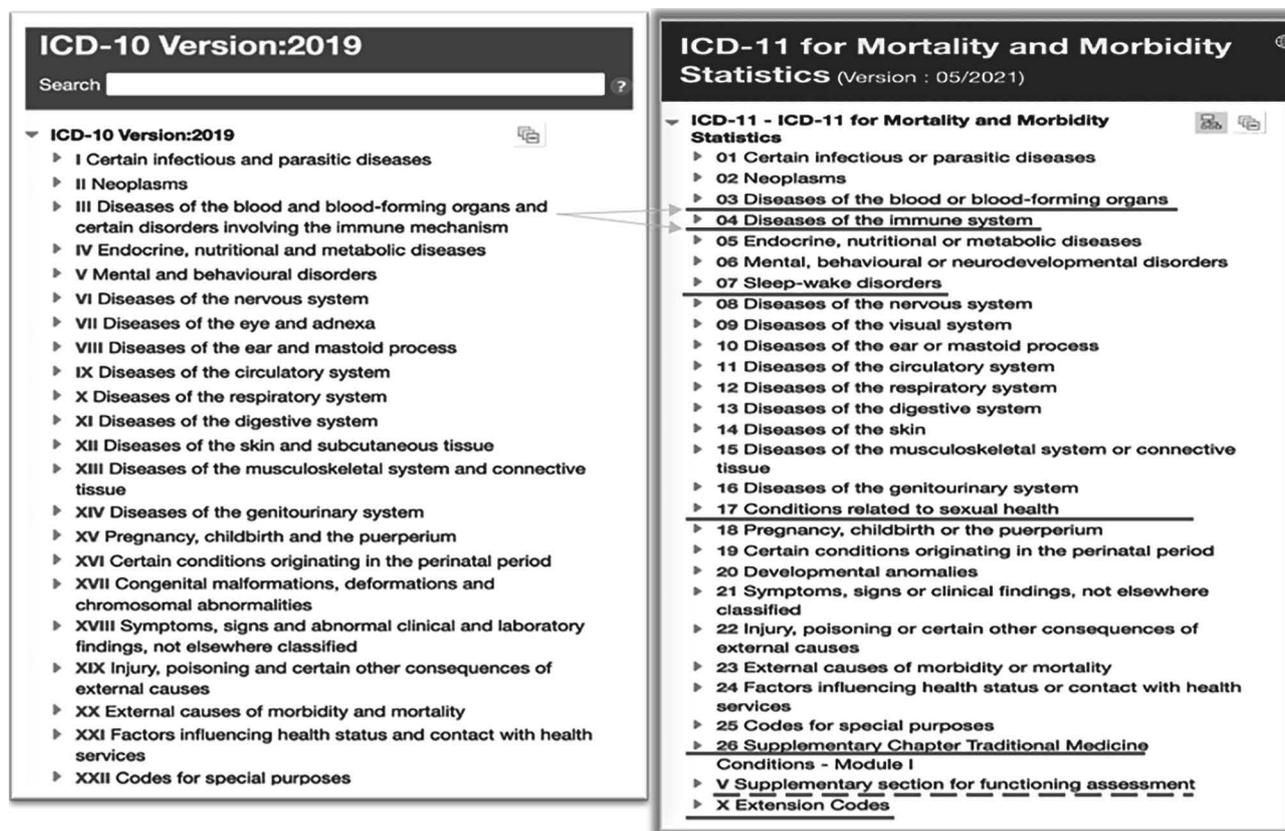


Figure 3. Chapters of ICD-10 and ICD-11.

(Compiled using the captured screens from the ICD-10 and ICD-11 browsers, with the new ICD-11 chapters underlined.)

or the circumstances for which the patient is receiving care at that particular time, or otherwise having some relevance to that person's care. Chapter XXII is dedicated to "Codes for Special Purposes," and the chapter's imperative functions can be seen during the pandemic (to be discussed in the following section 4.4).

ICD-11 maintained all ICD-10 chapters, while also having a vast enlargement of ICD-10. There are 28 chapters in ICD-11, of which 25 convey health conditions similar to past ICD versions, while one serves to identify external causes of morbidity and mortality, and another includes concepts of traditional medicine (World Health Organization 2022a, Sec. 3.2). As highlighted in Figure 3 above, chapters "03 Diseases of the blood or blood-forming organs" and "04 Diseases of the immune system" were split from a previous unified Chapter III of ICD-10, distinguishing differences in etiology, manifestations, and care (Harrison et al. 2021). New chapters for "07 Sleep-Wake Disorders" and "17 Conditions related to sexual health" consolidate more prominent topics since ICD-10. The totally fresh chapter "26 Supplementary Chapter Traditional Medicine Conditions" extends the classification to cover traditional medicine concepts, allowing the assessment of functioning and the optional recording of traditional medicine diagnoses.

There are two additional sections for optional use. The new unit V "Supplementary section for functioning assessment" addresses the need for some national modifications of ICD-10 for recording patient functioning. It is aligned with *International Classification of Functioning, Disability and Health* (ICF), another reference classification of the WHO Family of International Classifications Network. An addendum chapter X is called "Extension Codes" and codes starting with 'X' indicate an extension code. When linked to a stem code representing a clinical state, the extension codes add significant detail and allow for multidimensional coding. With its extension codes implemented, ICD-11 has the potential to improve precision and evidence-based health care worldwide (Drösler et al. 2021).

4.2 Notation system / code structure

The notation system of the classification is signified as "code structure" in ICD (World Health Organization 2022a, Sec. 1.2.4). The code structures of ICD-10 and ICD-11 are all significantly different from those used in previous versions and from each other. Figure 4 reveals their uniqueness, with examples from ICD-10 on the left side and ICD-11 on the right side.

ICD's chapter numbers are not parts in the ICD notations. ICD-10 uses Roman numerals (I-XII) as chapter numbers. ICD-11 chapter numbers are Arabic, which use a double-digit pattern for main chapters (01-26), except the supplemental chapters (V and X).

4.2.1 ICD-10 code structure

The numeric notation system used in ICD-9 and previous revisions was replaced by an alphanumeric code structure in ICD-10 that allows for more than double the size of the coding frame. "The main innovation in the proposals for the 10th revision was the use of an alphanumeric coding scheme of one letter followed by three numbers at the four-character level. This had the effect of more than doubling the size of the coding frame in comparison with the ninth revision and enabled the vast majority of chapters to be assigned a unique letter or group of letters, each capable of providing 100 three-character categories" (World Health Organization 2016 Vol.1, 15). The features are observable as the following (World Health Organization 2016, 2019b, 2022a, Sec.3.1):

1. The chapters are subdivided into homogeneous blocks of three-character categories. Each chapter contains sufficient three-character categories to cover its content; however, not all available codes are used, allowing space for future revision and expansion.
2. The first character of the ICD-10 code is a letter and does not relate to the chapter number. The letter may have been the same for two short chapters (e.g., Chapter VII (H00-H5) and Chapter VIII (H60-H95), or two letters may have been used for one long chapter (e.g., Chapter XIX (S00-T98)).
3. Of the 26 available letters, 25 had been used, the letter U being left vacant for future additions and changes and for possible interim classifications to solve difficulties arising at the national and international level between revisions. (Refer to "Figure 6. Emergency use ICD codes for COVID-19 disease outbreak" in Section 4.4 below, which confirms this important design.)
4. The range of categories is given in parentheses after each block title. [e.g., "F00-F09 Organic, including symptomatic, mental disorders" under "V Mental and behavioural disorders" showing in Figure 4). It is continued with "F10-F19," "F20-F29," and on].
5. ICD-10's stem code (category) is 3-characters [e.g., under the range of "F00-F09", there are "F00", "F01", "F02", "F03" (showing in Figure 4). It is continued with "F04" to "F09"].
6. The decimal point is always placed after the 3rd digit, followed by one digit.
7. Most of the three-character categories are subdivided by means of a fourth, numeric character (0-9) after a decimal point, allowing up to 10 subcategories (e.g., "F00.0" to "F00.9" under "F00" in Ch. V, showing in Figure 4). Where a three-character category is not subdivided, it is recommended that the letter 'X' be used to fill the fourth

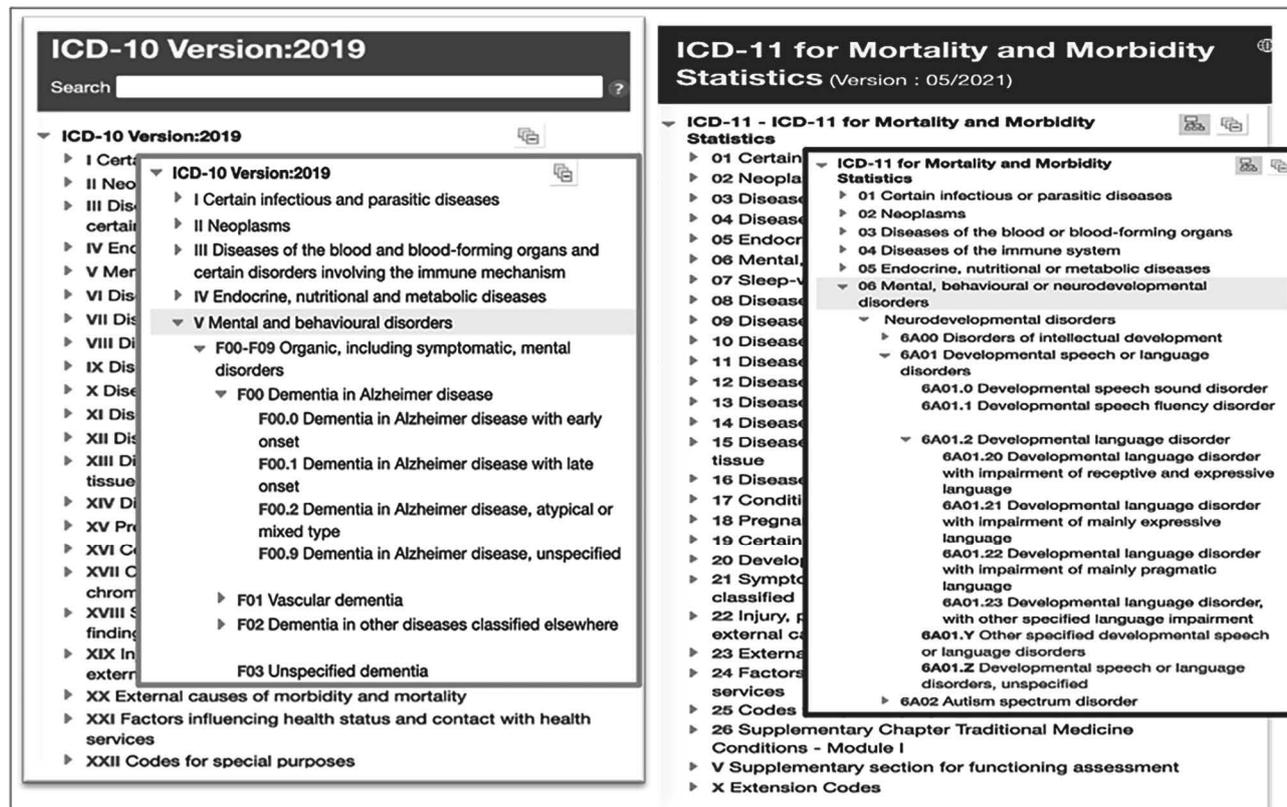


Figure 4. The code structures of ICD-10 and ICD-11.
(Compiled using the captured screens from the ICD-10 and ICD-11 browsers.)

position, so that the codes are of a standard length for data-processing.

8. The persistent terminal codes are: “.8” for a residual category (‘other’ conditions belonging to the three-character category); and “.9” for an unspecified category (e.g., “F00.9”, shown in Figure 4).

4.2.2 ICD-11 code structure

ICD-11 wants to ensure that data encoded with ICD-11 will be comparable with data encoded with ICD-10. Some patterns can be seen from these features, according to the *ICD-11 Reference Guide* section “1.2.4 General features” (World Health Organization 2022a, Sec.1.2.4.) and “Table 1: Major changes from ICD-10 to ICD-11, including rationale” in a section on ICD-11’s new conventions and terminology (Sec.3.1), with highlighted parts added, in the following examples.

1. Higher level entities in ICD-11 (called ‘blocks’) do not have category codes as they are not supposed to be used in coding, even though they may be used for reporting aggregated statistics. Blocks have their own URIs (e.g., the URI for Neoplasms is [02 Neoplasms]). Blocks may

also be referred to by block IDs with 11 characters long (e.g., “BlockL1-1A0”).

2. The Stem code (category) in ICD-11 is 4-characters, different from ICD-10 (3-characters). There are two levels of subcategories.
3. The codes of the ICD-11 are alphanumeric and cover the range from 1A00.00 to ZZ9Z.ZZ.

The letters ‘O’ and ‘I’ are omitted to prevent confusion with the numbers ‘0’ and ‘1’.

4. The first character of the stem code in ICD-11 always relates to the chapter’s order. It may be a number or a letter:
 - For the classes in Chapters 01 through 09, a first character of 1-9 is used (e.g., “6A00” and others under Ch. “06 Mental, behavioural or neurodevelopmental disorders”, as partially demonstrated in Figure 4).
 - For the classes in Chapters 10 through 26, the first character is a letter (e.g., “AB37” is in Chapter 10 and “SA10” in Chapter 26).
 - For the classes in the supplemental chapters V and X, the first character is the same as the one representing the chapter (e.g., “VD00” and “VW8Z” in Chapter V; “XS0T” and “XH1VJ3” in Chapter X). Codes starting with ‘X’ indicate an extension code.

5. In the second position of the stem code, there is a consistent letter to distinguish ICD-11 from the codes of ICD-10 (e.g., “6A00” under Chapter 06 (showing in Figure 4), “AB37” in Chapter 10, and “SA10” in Chapter 26).
6. The inclusion of a forced number at the third character position prevents the spelling of undesirable words.
7. A category (e.g., “6A01”) may have two levels of sub-categories (e.g., its sub-category “6A01.2” has further sub-categories, from “6A01.20” to “6A01.23”, shown in Figure 4).
8. The decimal point is always placed after the 4th digit.
9. The persistent terminal codes are: “Y” = a residual category (other specified); and “Z” = an unspecified category (e.g., “6A01.Y” and “6A01.Z”, shown in Figure 4, “6A0Y Other specified neurodevelopmental disorders”, and “6A0Z Neurodevelopmental disorders, unspecified” in Ch.6).
10. The linked diagnostic concepts, called a cluster, are supported by ICD-11 through post-coordination. This new feature enables linking core diagnostic concepts (i.e., stem + stem code concepts) when desired, and/or to add clinical concepts captured in extension codes to primary stem code concepts.

4.3 New features

In addition to the continuing content changes, which reflect knowledge and perspectives on diseases and their causes in each ICD version, changes in design and structure are noteworthy in ICD-11. The information framework is comprised of: 1) a semantic knowledge base (referred as the Foundation); 2) a biomedical ontology linked to the Foundation; and 3) classifications derived from the Foundation (Harrison et al. 2021). The Foundation component contains thousands of entities, which can be diseases, disorders, injuries, external causes, signs and symptoms, or reasons for encounter. Each entity is described by various properties, such as body structure, manifestation properties, causal properties, etc. Entities of the foundation become categories that are jointly exhaustive and mutually exclusive of each other (Grove and Jakob 2018). The adequate ontological design underlying ICD-11 for the uses and expressions of entities used in the knowledge base allows ICD-11 to represent the relevant domain of knowledge with a more straightforward way to enable incorporating new entities. For example, the SARS-CoV-2 virus, the COVID-19 disease, and various manifestations of the disease, which emerged after the release of ICD-11, were easily incorporated as new instances of relevant dimensions of the ICD-11 content model.

Three decades after the release of ICD-10, the world is now in the digital age. The significant features of ICD-11

include the improved ease and accuracy of coding, that requires less user training than ever before, and the availability of online and offline functioning (World Health Organization 2021-). ICD-11 comes with an implementation package that includes an online coding tool, an ICD-API that provides web services to enable remote programmatic access to ICD-11, enhanced capability to capture and combine clinically relevant characteristics of cases and integrated support for multiple languages, plus the *ICD-11 Implementation or Transition Guide* (World Health Organization 2019c). ICD-10 and ICD-11 mapping tables are provided correspondingly.

ICD-11 is combining the elements of classification and terminology and is designed to be linked to other terminologies that may provide additional detail or serve different purposes. Post-coordination is a notable new feature in ICD-11 that creates the ability to link core diagnostic concepts when desired, and/or to add clinical concepts captured in extension codes to primary stem code concepts. The linked diagnostic concepts are called a cluster (World Health Organization 2022a, Sec.1.2.4.5). In ICD-11, aligned with around 17,000 unique codes for the classification, more than 120,000 codable terms are now entirely digital, and the smart coding algorithm now interprets more than 1.6 million terms (World Health Organization 2022b). For example, when searching on COVID-19 related entries, a significant number of deprecated terms can be found, including those originally used a place name in naming a key variant of the virus that causes COVID-19. These deprecated terms lead to the formal classification entries.

For the first time in ICD, each of the ICD-11 entities has its Foundation URI (Uniform Resource Identifier), as demonstrated in Figure 5.

With the innovative use of the URIs, ICD-11 entities are fully language independent while having a specific place in a hierarchy of groups, categories, and narrower terms. Coding in ICD-11 can draw on statistical codes and on URIs (World Health Organization 2022a, Sec. 1.1.2). In this way, an international translation base facilitates translations or multilingual browsing. It will also eliminate the semantic conflicts and inconsistency issues when modifications are made to the classification after its worldwide implementation. In addition, since a statistical classification of diseases must be confined to a limited number of mutually exclusive categories that are able to encompass the complete range of morbid conditions, every disease or morbid condition must have a well-defined place in the list of categories. Nevertheless, there are residual categories for other and miscellaneous conditions that do not have their own unique category or cannot be allocated to the more specific categories. Thus, the ICD-11 URIs allow retention of such detail for future analysis as a key component in the taxonomy (Sec. 1.2.1).

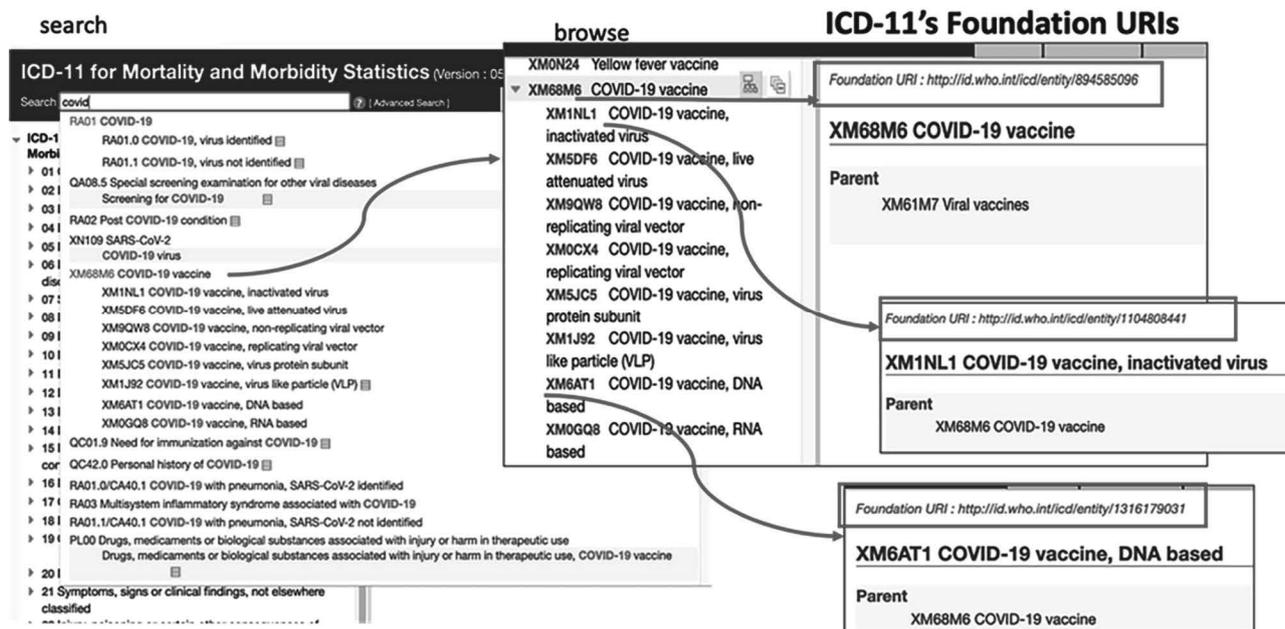


Figure 5. Examples of ICD-11's Foundation URIs.

(Compiled by the authors using the captured screens from the ICD-11 online searching and browsing website, with the URIs marked.)

As summarized by Grove and Jakob (2018), ICD-11's revision impetus enables ICD to

- Capture advances in health science and medical practice
- Make better use of the digital revolution
- Better address multiple topics, e.g., quality & safety, traditional medicine, etc.
- Address persistent major gaps in basic use for mortality statistics
- Improve morbidity statistics
- Easier use
- Manage national clinical modifications in more effective manner
- Improve integration of other classifications and terminologies
- Improve comparability of translations

4.4 Implications of the ICD during pandemics

During pandemics like COVID-19, it is more challenging to deal with information overload and semantic conflict; these challenges directly impact the whole domain of healthcare, which encompasses prevention, diagnosis, treatment, recovery or cures, illnesses, injuries, and other physical and mental impairments.

For the outbreak of a new viral disease, three very important names must be decided: 1) the disease, 2) the virus, and 3) the species (International Committee on Taxonomy

of Viruses Coronaviridae Study Group 2020). Establishing a name for a new disease provides a shared understanding for researchers and developers to discuss disease prevention, spread, transmissibility, severity, and treatment. The WHO guidelines ensure that the name does not refer to a geographical location, an animal, an individual, or a group of people, while still being pronounceable and related to the disease (World Health Organization 2015). An application of these guidelines can be seen in the naming of the disease, COVID-19, in February of 2020 (World Health Organization 2020a and 2020b) as well as in the labeling of key variants of the virus that causes COVID-19 using letters of the Greek alphabet in May 2021 (World Health Organization 2021b), in turn negating stigmatization and discrimination by avoiding naming according to the places where they are detected. ICD-10 established new emergency codes immediately after the WHO officially announced the name of the disease and provided guidance for COVID-19 coding in Feb. 2020. New classification codes have been added during the pandemic by both ICD-10 and ICD-11 (Figure 6).

ICD has been promoting international collaboration in the collection, processing, classification, and analysis of global health data, which allows for easy storage and retrieval of health information as well as sharing of health information among hospitals, regions, and countries. Without the ICD, each country or region would have to use its own disease classifications, which would most likely only be relevant for a specific area. As Bowker and Star (1999)

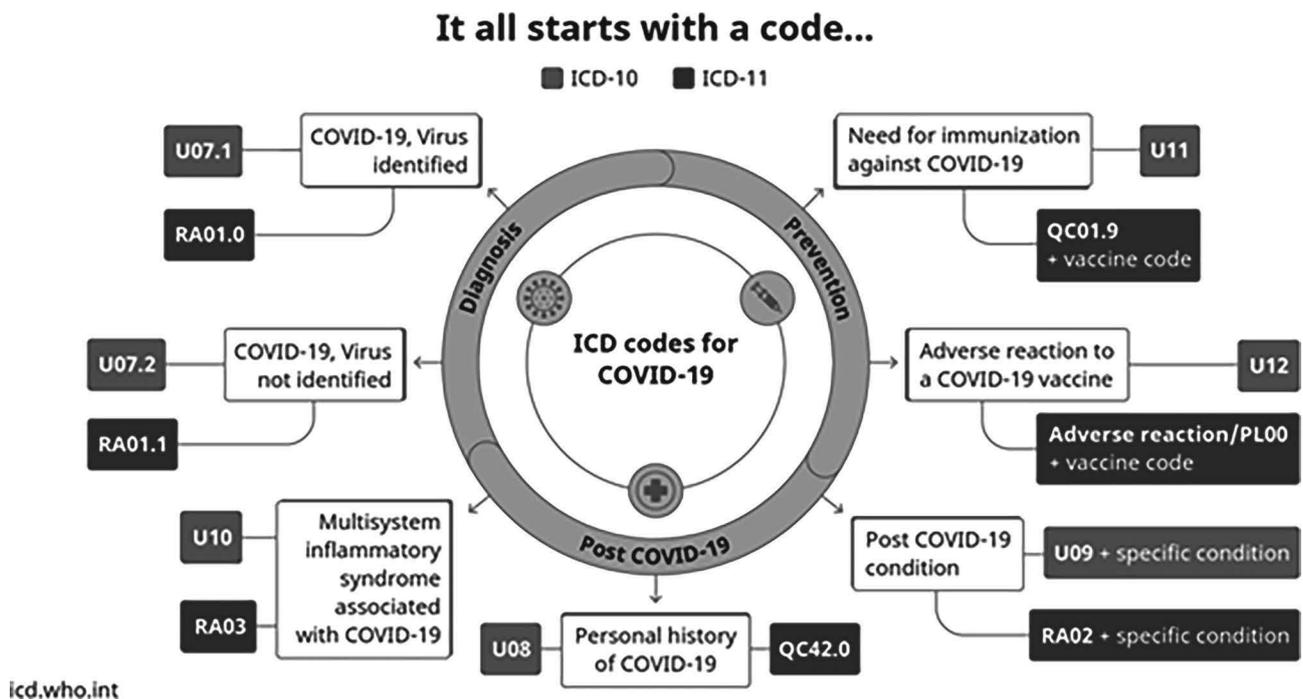


Figure 6: Emergency use ICD codes for COVID-19 disease outbreak.
(World Health Organization 2021c)

stated: “The ICD, for example, moves information across the globe, over decades, and across multiple conflicting medical belief and practice systems.” As reviewed by the authors in early 2020 (Zeng, Hong et al. 2020), ICD’s quick actions have been followed by other standardized health KOSs which have become even more critical to aid the frontline endeavors during devastating historic and worldwide events like the COVID-19 pandemic.

5.0 ICD’s KO practices: a short summary

Knowledge organization (KO) may be understood in a broad sense. The conceptual systems, the social fields, and the activity systems exist both inside and outside of KO, and each take place within different approaches and theories in all spheres of society. Therefore, the broader kinds of KOS (e.g., activity systems and scientific theories) are important for information professionals to study and understand (Hjørland 2016). ICD has not been commonly used or studied by information professionals (Cervone 2018); however, ICD’s utilization of the hierarchical classification structure, positional notations following syntactical rules, and methods in ensuring disambiguation, inclusion, consistency, and continuing extension, as revealed in the brief history of ICD’s development and implementations (refer to Section 3 and 4 above), can all be understood following classification theories and practices. This brief summary looks into eight of them.

1. “Nosology” is the term used in Sauvages’ *Nosologia Methodica*, which was credited as the first attempt to classify diseases systematically (Knibbs 1929). Nosology is the science of defining and classifying diseases. Over the past 200 years, medical science has been very successful in correlating obvious pathological conditions of the human body with patho-anatomical lesions, physiological disturbances, infectious agents (e.g., bacteria), genetic abnormalities, etc. (Severinsen 2001). Besides classifying concrete cases of disease-by-disease mechanisms or causal processes (“principle of disease mechanism”), other situations have been deeply discussed by philosophers and medical scientists, who have published their views on the fundamental principles behind definitions of diseases. These include modifying existing classification to conform with new information about the underlying mechanisms of diseases (e.g., mental illness), dealing with any still unknown causal structure and the complex causal structures behind diseases, investigating the similarities and differences between cases of disease, etc., as reviewed by Severinsen (2001), alongside a proposal of a pragmatic alternative and two “pragmatic principles.” There is also a “meta-nosology” (Livingstone-Banks 2018), which is a kind of study focusing on how diseases are defined and classified, what principles nosological practices are based on, the quality of the resulting medical taxonomy, and primarily whether/how diseases can be defined better than they are now.

2. “Classification” has been used in the title of ICD since revision #6 *International Statistical Classification of Dis-*

eases, Injuries, and Causes of Death. Originally, the revisions from #1 to #5 had the title *International List of Causes of Death*. The change of the title from a “list” to a “classification” in ICD-6 and its continual revising by the WHO through ICD-7 to ICD-11 reflects the formal identification and implementation of classificatory structures in knowledge organization. As Iyer Hemalata stated regarding the classical paradigm, “[t]he classical theory of concept formation is based on the assumption that the real world is structured in hierarchical groups that share common inherent properties” (Iyer 1995, 41). Aristotle, the founder of this view, developed a theory of classification in which all elements in a given class share at least one characteristic with all other members. Classes should be designed so membership of a class is given by a set of necessary and sufficient characteristics (Hjørland 2017, Sec.4.1).⁵ In the newest *ICD-11 Reference Guide*, the definition of “classification” is used according to the *ISO 17115:2020 Health Informatics - Representation of Categorical Structures Of Terminology (CatStructure)*: A classification is “an exhaustive set of mutually exclusive categories to aggregate data at a pre-prescribed level of specialization for a specific purpose” (World Health Organization 2022a, Sec. 1.1.2). The ICD has been developed as a practical, rather than a purely theoretical, classification in which there are several compromises between classification based on aetiology, anatomical site, circumstances of onset, or other criteria (Sec. 1.2.1).

3. As a statistical classification of diseases, the term “categories” was chosen by ICD regarding the features of ICD’s components, intending to facilitate the statistical study of disease phenomena, since it must be confined to a limited number of mutually exclusive categories and be able to encompass the complete range of morbid conditions (Sec. 1.2.1). The broad meaning of classification is the process of distinguishing and distributing kinds of “things” into different groups. Specific requirements or restrictions normally are seen in practice, such as that classes should be mutually exclusive and jointly exhaustive (Hjørland 2017). From the classical theory point of view, there are three assumptions (Smith and Medin 1981, 23-5): 1) Concept representation involves summary representation of the class; 2) The defining features for a category are both individually necessary and jointly sufficient to define the category; 3) Categories are nested, so that the subordinate categories possess all the features of the superordinate category. Additionally, categories could be formed based on family resemblances, rather than clearly marked definitions (Wittgenstein 1953, 328-9).

These norms are reflected in the ICD’s practices while the specific situations that health KOSs have faced are also considered. For instance, low frequency concepts tend to be grouped, but rare concepts may be individually classified if necessary (World Health Organization 2022a, Sec. 1.1.2). In

this continually revised classification, although the forms of revision are similar to others, including expansion, reduction and discontinuation, and relocation (Chan and Salaba 2016, 600), the workflow of ICD’s revision is dissimilar to customs in other classifications. Using ICD-11 as an example, the very large and diverse team (divided into multiple working groups) has involved over 300 specialists from many countries, with an additional 10,000 proposals from people all over the world (World Health Organization 2019c). More than 90 countries have been involved in production, reviews, testing or commenting (Grove and Jakob 2018). To determine whether an entity qualifies to become a unique category, ICD has been applying a set of measures, including epidemiological evidence, clinical evidence, granularity, continuity, and parsimony (World Health Organization 2022a, Sec. 1.1.2). This practical approach aims to retain the dual abilities to identify specific disease entities and to allow statistical presentation of data for broader groups, thus enabling the attainment of useful and understandable information.

4. The classification notation (code structure) designed for the various ICD revisions, particularly those revisions since ICD-6, revealed the hierarchical structure of the classification. ICD-6, released in 1948, employed three-digit Arabic numerals for class/subclasses and decimal points for subsections (refer to Figure 2 for a set of notations in ICD-6). It is noticeable that the 1940s witnessed the wide adoption of the *Dewey Decimal Classification* (DDC) in libraries. In a timeline provided by the Dewey Services [2017], the Library of Congress began to print Dewey numbers on catalog cards in 1930 and the Decimal Classification Committee, a forerunner to the present-day Dewey Decimal Classification Editorial Policy Committee, was established in 1937. By the 14th edition in 1942, the DDC index was over 1,900 pages in length and was published in two volumes, with around 31,000 entries (Comaromi 1976).

Notations are systems of written symbols that can be combined according to some set of syntactical rules to represent various meanings in a specialized domain. Its main function is mechanical control of concept ordering (Gnoli 2018). The design of the notation is considered to be one of the final steps in the compilation of a classification scheme (Hunter 2009, 73) or the third level of any classification scheme development process after the conceptual level (choosing the subjects and the way of structuring them) and the terminological level (Iyer 1995, 30-39). The commonly recognized norms of notation comprise uniqueness, simplicity, brevity, hospitality, flexibility, expressiveness, and retroactivity (Hunter 2009, 73-84). The distinctive features of the notational plane, when compared to the verbal plane, are uniqueness of the idea represented by an ordinal number and the total absence of homonyms and synonyms (Ranganathan 1967, 327-8; Gnoli 2018).

As a daily-used coding standard for health data exchanges in the digital age, the challenges ICD faces are much more significant than for bibliographic classifications and thesauri. The requirements for the background and experiences of the classification creators, the length of time in developing, testing, and finalizing any new version, and the official processes such as the formal approval and adoption by the WHO and individual countries, all demonstrate a whole new world that information professionals need to understand. Since ICD-10 substantial changes in both the main class notations and the rules used in the code structures are significantly observable, as shown in Section 4 of this paper. It is also noticeable that the ICD chapter's numbers are not functioning as the way that the common library classifications do, as they are not considered as the stem codes of the basic classes in a classification scheme.

5. An index is a kind of target document, which functions by providing access to information in or about some source documents in two ways: by deriving symbols from the source documents, or by assigning symbols about the source documents, thereby providing users access from a known order of symbols (e.g., A-Z) to an unknown place of information (Hjørland 2018). Starting with the ICD-2 English version, it contained an expanded alphabetical index, which showed the source of the items and gave the rubric numbers of both the detailed and abridged lists (U.S. Bureau of the Census 1911). In the digital age, the index-based search algorithm of the ICD-11 service interprets more than one million terms in ICD services. When using the unified Maintenance Platform of the WHO Family of International Classifications (WHO-FIC), the advanced search allows properties to be selected to include title, synonym, description, inclusion, exclusion, additional information, fully specified name, and narrower term.

6. Foundation URIs are officially implemented in ICD-11. A Uniform Resource Identifier (URI) is a compact string of characters for identifying an abstract or physical resource (Berners-Lee, Fielding and Masinter 1998). In ICD-11, each entity has a URI which is independent of the classification notation. The URIs can be used in a web browser or the expressions in different languages for an entity. In addition to ensuring the ICD feature of language-independence and maintaining the classification structure, this URI practice will enhance ICD's consistency and interoperability in the implementations, while allowing binding of any desired language to the elements of its foundation component (World Health Organization 2022a, Sec. 1.2.9).

7. ICD-11 is more than a revision of previous classifications within ICD; ICD-11 is claimed to be a classification and terminology. Aligning with its approximately 17,000 unique codes for the classification, more than 120,000 codable terms are now entirely digital (World Health Organization. 2022d). Preferred terms and deprecated terms

are available in the official WHO standard languages, accumulated from translations of previous revisions of the ICD, with trained translators checking the results. The entries, complemented by synonyms, can extend to other languages, and align with other ontologies (Harrison et al. 2021). Post-coordination in ICD-11 brings significant changes compared with previous revisions. The entire ICD meets the four important principles of vocabulary control, defined by *ANSI/NISO Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, in design and development: eliminating ambiguity, controlling synonyms, establishing relationships among terms where appropriate, and testing and validation of terms.

8. After more than 150 years long adventure, from a list to a comprehensive classification and terminology system, ICD-11 is within mainstream ontology-based design. The migration of its sibling classifications *International Classification of Functioning, Disability and Health* (ICF) and *International Classification of Health Interventions* (ICHI) (refer to Section 7.2 below) to a singular ontological infrastructure has enabled the full integration of terminology and classification in a common platform (World Health Organization 2022a, Sec. 1.1). "Its electronic architecture allows assignment of unique identifiers to any condition listed - independently whether the condition is grouped in a statistical class or whether it represents a class of its own" (Sec. 1.2.1). The database of ICD-11 is considered a semantic network (Harrison et al. 2021), with each entry characterizing a disease, syndrome, or health-related phenomenon in a way that not only is descriptive but also specifies its relationships with other entities.

6.0 Functions and applications of ICD

The purpose of the ICD is to allow for the systematic recording, analysis, interpretation, and comparison of mortality and morbidity data collected in different countries or regions at different times. The intended use cases can be seen from a broad range, such as causes of death, morbidity, epidemiology, case-mix (diagnosis-related group), quality and patient safety, primary care, functioning assessment, research, prevention, substance (medication) or device safety, specific surveillance like antimicrobial resistance, cancer registration, injury research, etc. As a standard, it is the key to ensuring semantic interoperability for clinical documentation, decision support, and guidelines or recommendations (World Health Organization 2022a, Sec. 1.1).

6.1 Diagnostic coding with the ICD codes

Diagnostic coding is a process to transform written descriptions of diseases, illnesses, or injuries into diagnostic codes.

From a medical information processing point of view, one of the important functions that physicians perform in their daily routine is encoding diagnoses using a standard vocabulary such as ICD to overcome the discrepancy between what physicians do and what they document (Hohnloser and Soltanian 1994). The basic coding steps in an electronic environment are different from that in a paper environment (World Health Organization 2022a, Sec. 2.6), with the workflow usually including:

- Analyzing and abstracting the diagnosis or assessment documented in the medical records.
- Extracting diagnostic statements from medical records.
- Entering the statement or term in an electronic coding tool or look up the term in the Alphabetical Index of an ICD codebook using the conventions, guidelines of the ICD coding manual (Proctor and Young-Adams 2014).
- Selecting the most appropriate matching term, or the one closest to what you are looking for from amongst the displayed options. Verifying the result in the tabular list of the online browser view or the paper codebook for exclusions, inclusions and notes given at the level of that category, its grouping levels and at the chapter level.

Appropriate use of the ICD is essential for an accurate translation of the medical record diagnostic statement into standardized codes, as demonstrated by the following two circumstances. 1) ICD includes a chapter classifying mental and behavioral disorders, started with ICD-6 in 1948 when the WHO began to lead the ICD revision and coordinate the statistical activities among the national committees of various countries, entering a new era of international cooperation for vital and health statistics. An international survey of 205 psychiatrists in 66 countries found that ICD-10 was frequently used for clinical diagnosis and coding in mental health (Mezzich 2002). Phull (2012) presented a learning guide to use ICD-10 for psychiatry practice of assessment, diagnosis, and management, which is very helpful for anyone interested in learning about diagnostic coding in psychiatry, particularly medical students, psychiatry trainees, and mental health professionals. 2) Since the field of bereavement research and care has been at a tipping point in recent years, ICD-11 introduced prolonged grief disorder (PGD) as a diagnostic category on the international stage. The current definition of PGD for the ICD-11 is the latest iteration following nearly 30 years of research investigating different diagnostic definitions. Previously a disorder of grief has been described and researched under different names, which has led to some confusion amongst clinicians and researchers particularly during the COVID-19 pandemic. The multiple definitions and disorder criteria can lead to different diagnostic thresholds and estimates of incidence. The newest ICD-11 definition of PGD deviates

from previous definitions as it seeks to align with the WHO's remit for improved clinical utility and global applicability of mental disorders (Killikelly et al. 2021; Eisma et al. 2021).

During the COVID-19 pandemic, the use of ICD codes was critical. Within two years, the special portal "COVID-19 Global Literature on Coronavirus Disease" of the WHO (<https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/>, mainly from MEDLINE and EMBASE, with over 20 resources) reveals more than 750 articles when searching "ICD" as of March 29, 2022. Nearly one sixth of them (127 articles) have the main subject "COVID-19". Other main subjects of the articles include Pandemics (21), Pneumonia, Viral (15), Hospitalization (15), Coronavirus Infections (14), and SARS-CoV-2 (10). For the articles focused on ICD (13 articles), the types of studies are categorized as Diagnostic study (7), Prognostic study (5), Risk factors (4), Etiology study (2), Incidence study (2), Observational study (2), Clinical Practice Guide (1), Qualitative research (1), and Screening study (1). Half of these articles concentrated on the validity of ICD-10 codes and accuracy.

Several of these ICD-focused articles brought the processes of auto-recording, predicting, generating data analysis and reports using ICD. The articles also presented new efforts in the development of ontologies and applications based on ICD-11. In fact, automated ICD coding has been investigated and developed in the health information community. Since ICD is getting more and more complex, even well-trained medical coders often find accuracy in diagnostic coding challenging with the manual coding process. This is especially true when new ICD codes are supposed to be used immediately and accurately, e.g., when the pandemic started in early 2020, when there was a lack of initial awareness or familiarity with ICD-10 coding for COVID-19 (Bhatt et al. 2021). Taking advantage of ICD being a diagnostic classification with a hierarchical tree of codes, the diagnosis code assignment modeling would be improved when the hierarchical nature of ICD codes is leveraged. Predicting ICD codes according to the discharge summary content is an example of large-scale modeling applied to a routine healthcare task (Perotte et al. 2014).

6.2 Clinical documentation with ICD

Clinical documentation includes all clinical notes and reports that relate to patient care. It has numerous benefits for the immediate care of the patient and more advantages for medical practice at large (Mamykina et al. 2012). Clinical documentation consists of a process in which healthcare providers record the observations, impressions, plans, and other activities arising from episodes of patient care, and generally occurs with each interaction between patients and the healthcare system. This produces an objective record of

a patient's medical history, physical findings, medical reasoning, and clinical problem solving (Rosenbloom et al. 2010). Since a diagnosis is in line with the collection of information about the history, symptoms, and signs of a case, a critical component of clinical problem solving involves gathering and organizing this data in patient records (Patel et al. 2000). The sources of diagnostic statements are presented with various clinical documents, including encounter forms, history and physical reports, progress and treatment notes, discharge summaries, procedure and operative reports, radiology reports, pathology reports, laboratory reports, etc. The diagnoses and procedures listed in encounter forms, clinical notes, and reports are usually encoded with ICD codes for standardized patient records.

Considering that not every problem or reason for conducting health services can be categorized in diseases and injuries, ICD correspondingly incorporates a wide variety of signs, symptoms, abnormal findings, complaints, and social factors that represent the content from health-related records (World Health Organization 2016, Vol.2, 2). The ICD can therefore be used to classify data recorded under headings such as "Cause of death", "Diagnosis", "Reason for admission", "Conditions treated", "Additional diagnoses", "Risk factors", and "Reason for consultation", which appear on a wide variety of health records and documents, from which statistics are derived for treatment, prevention, or patient safety (World Health Organization 2022a, Sec. 1.1).

Thus, the medical conditions, health problems, diagnoses, and procedures in these clinical notes and reports are often encoded with ICD codes. The use of the ICD facilitates accurate clinical documentation and efficient processing of claims for the disease or injury for which a patient was treated, among many benefits.

6.3 Communication using the ICD codes in Electronic Health Records

Electronic Health Records (EHR), also referred to as Electronic Medical Records (EMR), support the documentation of clinical practice at the point of care. The EHR is a rich source of information on human diseases, but the information is variably structured, fragmented, curated using different coding systems, and collected for purposes other than medical research (Denaxas et al. 2019). Schiff et al. (2010) emphasized that electronic clinical documentation with EHR systems may enhance communication among clinicians, reduce error rates, improve documentation efficiency, and help prevent diagnostic errors.

Standardized KOSs are one of the keys to ensure interoperability among EHR systems (refer to Section 7.1). Among the KOSs implemented in EHR, the ICD is mainly used to normalize the diagnostic process, health problem lists, and

medical procedures. A study by Keck et al. (2014) using ICD-9 in an EHR-based influenza-like illness (ILI) surveillance system, was developed during the 2009 influenza A (H1N1) virus pandemic for the American Indian and Alaska Native population. It showed strong evidence that the ILI surveillance had increased accuracy and timeliness. The implementation of ICD-9 provided useful information for decision-makers, as indicated by the EHR, and offered the opportunity to transform public health surveillance. Other cases using ICD within EHR in the USA revealed that surveillance using ICD-9 codes might be reliable to determine the burden of a chronic hepatitis B virus (HBV) infection and improved reporting by state and local health departments (Mahajan et al. 2013). A different, unique study by Blosnich et al. (2018) showed that transgender individuals were vulnerable to negative health risks and outcomes. Still, research remains limited because data sources such as the EHR lack standardized collection of gender identity information. Using ICD-defined transgender status can facilitate health services research when self-identified gender identity data are unavailable in EHR.

A typical workflow facilitated by a tool is demonstrated by Horsky, Drucker, and Ramelson's 2017 study. When a clinician starts to add a clinical term to a problem list by typing one or more words into a free-text query field, an ICD-10 code lookup pops up in the EHR. A search engine within the EHR then returns a list of ICD-10 codes and descriptions and ranks them according to their relevance to the search string for complete word or partial word matching. If there are no exactly matched words found with the query, clinicians may find it challenging to choose "close enough" codes. In addition, this coding process requires accurate documentation of clinical care, made possible with the ICD-10 system.

7.0 ICD and other health KOSs

Standardized health KOSs are playing an increasingly larger and more important role in healthcare information systems to facilitate data normalization, which is a fundamental requirement for any subsequent data analysis, information management, and decision-making. ICD and other standardized KOSs allow the world to exchange and share data consistently and in a standard way, between institutions, across regions and countries, and over time. In addition, they facilitate the collection and storage of data for analysis and evidence-based decision-making.

The world has witnessed the growing development and usage of KOSs in healthcare information systems. In addition to the ICD and the WHO Family of International Classifications Network, hundreds of biomedical KOS vocabularies can be found in biomedical KOS repositories.

7.1 ICD and other common health KOSs

With the realities of everyday healthcare, various KOS standards are needed for different processes. This illustration (Figure 7) displays their major usage in EHR, according to practices in the USA. As shown in the figure, ICD and SNOMED CT (*Systematized Nomenclature of Medicine - Clinical Terms*) are mainly used to normalize the diagnostic process and health problem list. SNOMED CT is the most comprehensive clinical terminology in use around the world and could be used for almost all medical terminology coding processes. There are others that have been developed to mainly encode medical lab tests and clinical observations (e.g., *Logical Observation Identifiers Names and Codes* [LOINC]), medical procedures (e.g., *Current Procedural Terminology* [CPT]), and names for clinical drugs (e.g., RxNorm). They have a crucial role in healthcare information systems to support data normalization and semantic linking and sharing.

The Health Level Seven (HL7) messaging (“Level Seven” signifies the seventh level [the application level] of the ISO seven-layer communications model for Open Systems Interconnection [OSI]) is the workhorse of electronic data exchange in the clinical domain and arguably the most widely implemented standard for healthcare in the world. This messaging standard allows the exchange of clinical data between systems. It is designed to support a central patient care system as well as a more distributed environment where data resides in departmental systems (HL7 International 2007-2022). HL7 requires the use of standardized terminologies to represent health data. Besides developing its own standardized code sets to identify administrative data such as gender code, data type, and status codes, HL7 has employed existing standardized health KOSs to support the disambiguation of health information exchange. These

KOS standards have an essential role in healthcare information systems to support data normalization. Among them, ICD, CPT, SNOMED CT, LOINC, and RxNorm are the most widely used KOS standards in EHR and health information exchange (HIE). (Refer to Appendix for a list of health KOS).

Going beyond those standards listed in Figure 7, which have evolved solely from practices in the USA, more than 200 biomedical KOS vocabularies are included in the Unified Medical Language System (UMLS) Metathesaurus, which is a large, multi-purpose, and multilingual biomedical vocabulary database. It is organized by concept, with links to similar names for the same concept from over 150 different KOS sources (distinct source families by language) across the world (mainly in English, about 70 non-English, as of June 2022) (U.S. National Library of Medicine 2022). Major categories in the Metathesaurus (U.S. National Library of Medicine 2016-) embrace diagnosis and clinical observations, procedures & supplies, diseases, comprehensive vocabularies/thesauri, and other categories. In addition, some vocabularies fall into more than one category.

In the UMLS Metathesaurus, ICD-10 is featured under the “Diseases” category in the source vocabulary (U.S. National Library of Medicine 2016-). Multiple members of the ICD family can be found, incorporating different languages and coverages, for example, *ICD-10 German*, *ICD-10, American English Equivalents*, *ICD-10, Australian Modification*, *ICD-10, Australian Modification, American English Equivalents*, *ICD-10, Dutch Translation*, etc. In addition, several specific ones listed in the WHO family of KOSs (refer to Section 7.2) are included. The *International Classification of Primary Care* (ICPC) alone has 14 languages. There are products derived from the ICD, including *ICD-10 Clinical Modification* (ICD-10-CM) which are widely used in the USA and other countries.

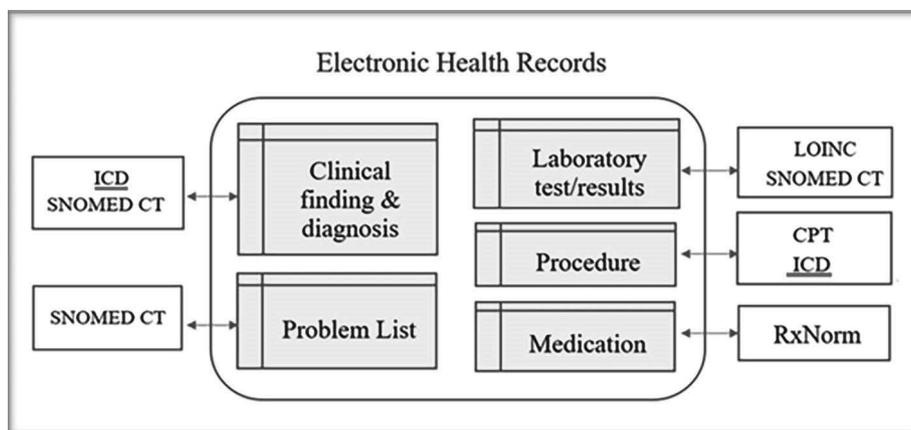


Figure 7. Main usages of different standardized health KOSs in EHR, according to practices in the USA. (Image created by authors.)

Another terminology repository, BioPortal, the world's most comprehensive repository of biomedical ontologies developed by the National Center for Biomedical Ontology (2005-), contains 868 ontologies as of May 6, 2021, and 1,005 as of Aug. 10, 2022. These numbers indicate the need for standardized KOS vocabularies in diverse circumstances and communities alongside cross-vocabulary mappings that facilitate data and knowledge as semantically interoperable across knowledge sources that are developed independently to meet diverse needs.

More than a hundred vocabularies in BioPortal are mapped to ICD-10; nearly 20 of them have mappings with over a thousand classes. The *Radiology Gamuts Ontology*, for example, mapped 18,501 ontological classes with ICD-10. (Refer to ICD-10 entry at BioPortal: <https://biportal.bioontology.org/ontologies/ICD10>). As gender and sexual orientation continue to represent crucial areas of medical practice and research with evolving terminology, a *Gender, Sex, and Sexual Orientation (GSSO)* ontology released its second version to address this gap by providing a centralized data resource after evaluating and mapping with ICD and other major KOSs (Kronk and Dexheimer 2020). Many ontologies have been timely released and updated during the COVID-19 pandemic (e.g., *The COVID-19 Infectious Disease Ontology*; *An Ontology for Collection and Analysis of COviD-19 Data*; *COVID-19 Ontology*; and *COVID-19 Ontology in Pattern Medicine*). They are all mapped to ICD-10, as shown in BioPortal. Countless specific studies on the sufficiency and the needed extensions built on ICD-9 and ICD-10 have been reported, such as in the *Journal of the American Medical Informatics Association* and the Annual Symposium of the American Medical Informatics Association.

7.2 ICD in the WHO Family of International Classifications (WHO-FIC) Network

Led by their mission to improve health through the ongoing development, maintenance, and promotion of an integrated suite of reference health classifications and related products that produce information of value and utility across the world, WHO has a Family of International Classifications (WHO-FIC) Network. "The WHO-FIC provides a standardized common language for coding a wide range of information about health and wellbeing, and about interventions delivered to individuals and populations. The classifications and terminologies have been developed based on sound scientific principles, ensuring that the classifications are internationally and culturally appropriate and can convey information in an unambiguous way across languages, cultures, settings and borders" (World Health Organization 2021d). The classifications are categorized into three groups: Reference classifications, Derived classifications, and Related clas-

sifications (Madden, Sykes and Ustun [2007]; World Health Organization 2021d). Among them, ICD is the most important one, as highlighted in Figure 8.

The reference classifications (listed in the center column in Figure 8) are developed for the classifications of diseases (ICD), functioning, disability, and health (ICF), and health interventions (ICHI) for international reporting, and cover the main parameters of health. ICD and ICF have been developed to complement each other to capture the full picture of an individual's health status since they offer different coverage of health terms. While the ICD classifies disease and other health conditions for diagnostic coding, ICF, as spelled out in the title, defines domains of functioning and disability in terms of body functions and structures or activities. ICHI (work in progress) has been developed to provide a common language tool for describing, reporting, and analyzing health interventions for statistical purposes. The classification is built around three axes: Target (the entity on which the Action is carried out), Action (a deed done by an actor to a Target) and Means (the processes and methods by which the Action is carried out) (World Health Organization 2007-). The WHO is responsible for these primary reference classifications. They may be used together, for example, for capturing information on health interventions relevant to people with disabilities, in which the ICHI is used to record investigative and diagnostic interventions conducted along with the therapeutic and supportive interventions subsequently delivered; ICF is used to describe the person's functioning and need for environmental facilitators; and ICD is used to record a diagnosis (Fortune, Madden and Almborg 2018; Manchaiah, Swanepoel and Fortune 2020). A WHO-FIC Maintenance Platform (<https://icd.who.int/dev11/>) provides a unified search and browser for unreleased, work in progress versions ICD-11, ICF, ICHI, and several other classifications. The audience for this site is the maintainers, contributors, and translators of the classification.

A set of the derived classifications (right hand column in Figure 8) are extensions of the core reference classifications that allow for further depth within a particular area or setting. They have been developed based upon the ICD by adopting its structure and categories and providing additional details beyond what is provided by the ICD, involving collaborations with specialist user groups. The focus currently is on the classifications of: the diseases for oncology (ICD-O-3), mental and behavioral disorders, neurology (ICD-NA), dentistry and stomatology (ICD-DA), and reporting of causes of death in low-resource settings (ICD-10-SMoL). There are no semantic conflicts between ICD and these derived classifications since they define diseases and disorders consistently while covering more detailed information in a specific area such as oncology and mental health.

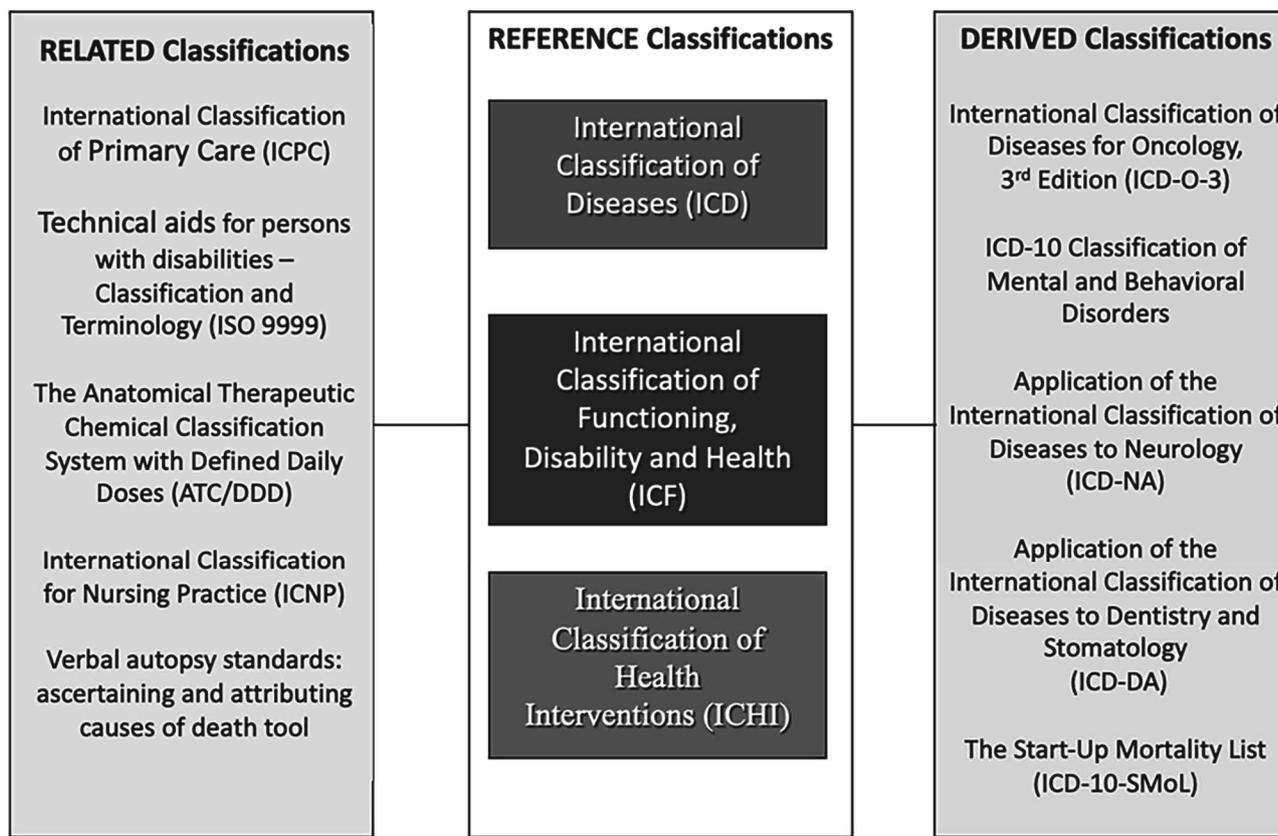


Figure 8. WHO Family of International Classifications (WHO-FIC) Network.
 (Image modified based on Madden, Sykes, and Ustun [2007] and World Health Organization 2021d)

The related classifications (left column in Figure 8) are those that partially refer to or are associated with the reference classification at specific levels of structure (Madden, Sykes and Ustun [2007]). They are included in the WHO-FIC to describe important aspects of health care and functioning of the health system, including primary care, technical aids for persons with disabilities, nursing practices, and other standards that complement reference and derived classifications (World Health Organization 2021d). One of the classifications listed in this group, *International Classification of External Causes of Injury* (ICECI), has informed the redesign of the relevant chapter of ICD 11th revision. Hence the different elements of ICECI have been included as extension codes in the ICD-11.

8.0 ICD in health informatics

Parallel to the research articles published in medical journals, health informatics journals have accumulated cross-domain research reports, since clinical reports encoded with ICD codes are good resources for scientists to perform further analyses and develop advanced approaches in the digital age. In 1998, an article entitled “Desiderata for controlled

medical vocabularies in the twenty-first century” expressed the desiderata (Latin: “things desired”) of standard medical KOSs. Builders of medical informatics applications need controlled medical vocabularies to support their applications, and it is to their advantage to use available standards with multiple purposes including capturing clinical findings, natural language processing, indexing medical records, indexing medical literature, and representing medical knowledge (Cimino 1998). In 2017, another study assessing content and trends in medical informatics publications over the past two decades discovered a set of continuously growing core and emerging topics that medical informaticians have been dealing with in collaboration with professionals in healthcare and other technological fields (Wang et al.). As delivered by selected publications from *Journal of the American Medical Informatics Association*, *Journal of Biomedical Informatics*, *Biomedical Informatics*, *International Journal of Medical Informatics*, the series *Studies in Health Technology and Informatics*, and other journals covering health informatics, it is obvious that researchers have developed and implemented various information technologies using particular clinical documents and data while addressing various issues, as shown in the above sections as well as this special

section. The multiple dimensions of research in line with the clinical reports encoded with ICD can be seen in diverse topics and methods, as demonstrated below with the major categories across these research papers.

8.1 Information technology development and application

The assignment of ICD codes to health-related documents has become a focus of academic research. Numerous studies have developed the process of ICD coding from manual to automated work, as summarized by Yan et al. (2022) for the history of automated ICD coding research (Figure 9). The milestones can be seen in three development stages: Stage 1, Rule-based, Stage 2, Traditional Machine Learning-based, and Stage 3, Neural Network-based.

The application of artificial intelligence (AI) and machine learning (ML), along with natural language processing (NLP), may automatically extract information from clinical reports and reduce delays in report processing. However, the complicated content in clinical reports can be difficult for machines to process due to its high variability, including misspellings and missing punctuation, clinical diagnoses interspersed with complex explanations, and different terminology to represent the same disease and condition. Developing an automated solution with high accuracy and consistency across a wide selection of reports is therefore challenging. Use of standard ICD codes with their hierarchical structure is expected to significantly improve the accuracy of extracted health information. Considering the design, implementation, and validation of a novel deep learning (DL) approach, in line with the hierarchical attention networks (HANs), Gao

et al. (2018) performed automate information extraction from unstructured clinical pathology reports. The *International Classification of Diseases for Oncology* (ICD-O-3) topographical codes were employed for information extraction from 942 deidentified pathology reports that matched to 12 ICD-O-3 topography codes corresponding to 7 breast and 5 lung primary sites. Two cancer pathology report classification tasks demonstrated that HANs could effectively capture the primary information regarding a patient's tumor type, grade, and location from unstructured text.

Brown and Sönksen (2000) used a semantic terminological model for clinical findings and measured the strength of agreement between the concepts and records retrieved from a computerized patient database of 106,000 patient problem record entries containing 2,625 unique concepts. The performance of the model was compared with the semantic, free-text, ICD-10, and hierarchic data retrieval against a gold standard in response to a battery of 47 clinical questions. The use of a semantic terminological model in this test scenario provided an effective framework for representing clinical finding concepts and their relationships, which improved information retrieval from a patient database in response to clinically relevant questions.

Increasing use of EHR provides a new way to improve care for patients with chronic pain and other conditions. Tian, Zlateva and Anderson (2013) analyzed data from the eClinicalWorks EHR system showed that patients' demographics, clinic utilization, patient-reported pain scores, opioid prescription records, diagnostic codes, laboratory results, and referrals played an important role in determining an accurate and reliable method for identifying patients with chronic pain. The ICD codes from patient encounter

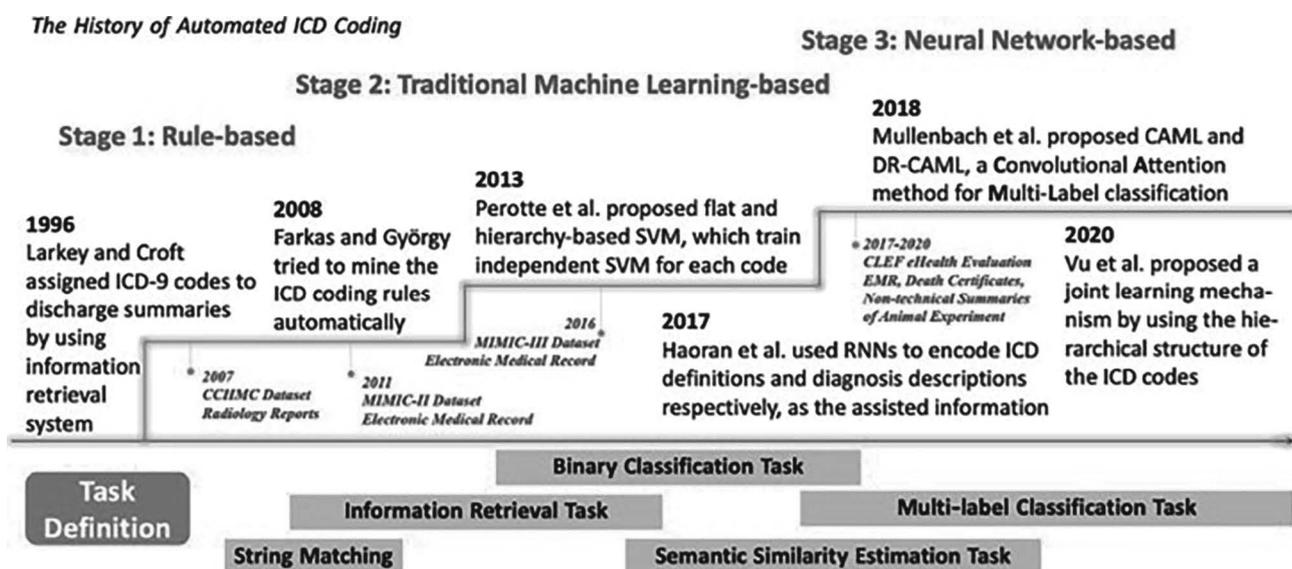


Figure 9. Developmental history of the automated ICD coding task (Yan et al. 2022).

diagnoses and problem list entries were used to develop an algorithm to accurately identify chronic pain. By developing the algorithm using critical data elements such as ICD diagnosis and procedure codes from EHR, this study enhanced the capability to identify chronic pain in order to address this public health challenge. Another study by Perotte et al. (2014) presented an automated diagnosis coding model and novel evaluation metrics based on the discharge summaries from the publicly available Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) repository. It reflected the distances among gold-standard and predicted codes and their locations in the ICD-9 tree.

Frequent pattern mining of EHR data is critical to identifying potential associations and correlations in EHR data that researchers may not consider or that may have otherwise gone unnoticed. Using a case study regarding characterizing pediatric asthma, Campbell, Bass and Masino (2020) introduced a temporal condition pattern mining methodology to address the sparse nature of coded condition concept utilization in EHR data after analyzing a large and complex EHR dataset for temporal condition patterns. The SPADE (Sequential Pattern Discovery using Equivalence classes) algorithm was applied to two datasets: one dataset with diagnoses coded using ICD concepts, and another dataset with the ICD codes mapped to their corresponding expanded diagnostic clusters. The analysis revealed strong associations between asthma and several comorbidities and temporal condition patterns, which can be used as hypotheses to explore causality in future pediatric asthma research. The methodology can be applied to identifying temporal patterns in EHR data to investigate conditions and research objectives in numerous contexts outside pediatric asthma.

8.2 Health and clinical data analysis

Health data are needed to measure performance, improve program decisions, and increase accountability, inform health system financing and resource allocation, and evidence-based practice. Health data are also used in guidelines and decision support, monitoring of health outcomes and access to services, and the assessment of the health of individuals and populations (World Health Organization 2021d). ICD in health and clinical data analysis are reported in a wide range of informatics discoveries. These studies embrace health data mining with ICD codes in certain types of clinical population or disease cases, diverse drug events, specific epidemics, emerging patterns of inequity, phenotyping performance of EHR components, diagnosis code evaluation, usage of ICD codes, alternative privacy protection, etc. In research by Wei et al. (2016), the phenotyping performance of three major EHR components (ICD diagnosis codes, primary notes, and specific medications) was evaluated using de-identified Vanderbilt EHR data with ten dis-

eases: atrial fibrillation, Alzheimer's disease, breast cancer, gout, human immunodeficiency virus infection, multiple sclerosis, Parkinson's disease, rheumatoid arthritis, and types 1 and 2 diabetes mellitus. For each disease, patients were classified into seven categories in line with the presence of evidence in diagnosis codes, primary notes, and specific medications. Review results were used to estimate the positive predictive value (PPV), sensitivity, and F-score for each EHR component alone and in combination with one another. It was found that multiple EHR components provided a more consistent and higher performance than a single EHR constituent for the selected phenotypes.

Since ICD codes are important components of national health data sources and EHR systems, evaluation, transformation, and usability of ICD coded data have been documented based on substantial data analyses and experiments. Many research projects have been reported while as many specific articles demonstrate how the data are used. Boslett et al. (2019) performed a secondary analysis of the 2014-2016 Multiple Cause of Death data from the CDC. The team combined various national data sources (mortality, demographic, economic, and geographic) in a multi-method analysis of missing drug classification in the overdose mortality records (as defined using ICD-10 code T50.9 on death certificates). Because the mortality data is a key data source used in measuring the scale of the opioid epidemic, it is indispensable to understand the individual and community inequities underlying the missing data on specific drug involvements. The findings from analyzing opioid-related data with ICD codes demonstrated that knowledge of these inequities could enhance understanding of the opioid crisis and inform data-driven interventions and policies to support public health and epidemiological efforts (Boslett et al. 2019). An evaluation study by Wiley et al. (2013) concentrated on the validity, usage, and potential research applications for ICD tobacco codes in clinical populations using data on cancer cases and cancer-free controls from Vanderbilt's biorepository, BioVU. The utility of ICD-9 tobacco use codes was evaluated to identify ever-smokers in general and high smoking prevalence (lung cancer) clinic populations. The potential biases in documentation were assessed while a temporal analysis relating transitions between smoking codes to smoking cessation attempts was performed. The study team examined the suitability of these codes for use in genetic association analyses and found that ICD-9 tobacco use codes could identify smokers in a general clinic population, there was little evidence of documentation bias, and frequency of code transitions between 'current' and 'former' tobacco use was significantly correlated with initial success at smoking cessation. Furthermore, these ICD codes were suitable for adjustment of smoking status in genetic studies utilizing EHR with some limitations (Wiley et al. 2013). More ICD codes were involved in

adverse drug event data mining. For example, Hohl et al. (2014) used a comprehensive set of ICD-10 codes along with coding algorithms to identify adverse drug events in administrative health data that was linked with medication dispensing and represented a rich source of health information. Adverse drug event data from this source offered accessible and standardized population-level data over long periods of time, enabling analysis of time trends, prescribing patterns, and comparisons across healthcare settings.

There is a significant interest in leveraging the electronic medical record (EMR) to conduct genome-wide association studies. A biorepository of DNA and plasma was created, and demographic data and laboratory values were extracted from the EMR. Medication use and smoking status were established by natural language processing of clinical notes. Other risk factors and comorbidities were ascertained based on ICD-9-CM codes, medication use, and laboratory data. This study demonstrates the use of the EMR to ascertain phenocopies, phenotype heterogeneity, and relevant covariates to enable genome-wide association studies of peripheral arterial disease. Biorepositories linked to EMR may provide a relatively efficient means of conducting GWAS (Kullo et al. 2010). On the other hand, the issue of participants' privacy was brought up by a study focused on the disclosure of diagnosis codes. The researchers examined whether released data could be linked with identified clinical records that were accessible via various resources to jeopardize patients' anonymity or the ability of popular privacy protection methodologies to prevent such an attack. By experimentally evaluating the re-identification risk of a de-identified sample of Vanderbilt's patient records involved in a genome-wide association study and measuring the level of protection from re-identification and data utility provided by suppression and generalization, the study (Loukides, Denny and Malin 2010) concluded that popular privacy protection methods were inadequate to deliver a sufficiently protected and useful result when sharing data derived from complex clinical systems. The development of alternative privacy protection models was thus required.

8.3 Enhancing quality, interoperability, comparability, and reusability

The nature of everyday healthcare has been leading various KOS vocabularies to be developed and used for different processes, including various KOSs in the WHO-FIC network, the multilingual versions of ICD, the national clinical modifications, the correlated ICD derivations, and other common health KOSs (refer to Section 7 and Appendix). Research and development related to this area commonly involve multiple KOSs, as demonstrated in the following sub-sections.

8.3.1 Comparative studies involving ICD and multiple health KOSs

Comparing different KOSs has been a common method used by researchers. In an evaluation of clinical coding schemes involving medical as well as ample nursing terminology, James Campbell and the research team (Campbell et al. 1997) considered criteria of completeness, taxonomy, mapping, definitions, and clarity. The authors assembled 1,929 source concept records from a variety of clinical information taken from four medical centers across the United States. Classification system features were compared based on completeness, clarity (coding duplication rate), mapping (fraction of candidate concepts with administrative mapping), compositional, synonyms, attributes and uncertainty, taxonomy (# of first-degree relatives per concept), meaningless identifiers, unique identifiers, definitions (fraction of concepts with definitions provided), language independence, and syntax/grammar. The rich data collected and analyzed points out important priorities for classification scheme developers to pursue in order to improve their products.

ISO 1828:2012 Health informatics - Categorical structure for terminological systems of surgical procedures specifies the minimal characteristics of a categorical structure for terminological systems of surgical procedures and the minimal domain constraints. One goal of the standard is to support comparisons between existing classification systems of surgical procedures. It aims to support interoperability, comparability, and the exchange of meaningful information on surgical procedures independently of language, insofar as the significant differences are specified by the system. It is intended to be used as an integrated part of computer-based applications and for electronic health care records (ISO 1828:2012). By comparing several existing national and international classification systems of surgical procedures with the CEN/ISO 1828 Categorical Structure, a collaborative research and development group (Rodrigues et al. 2011) reported the findings and pointed out that such a comparison of the different existing classification systems based on this ontology framework presented a step towards increased interoperability between biomedical terminologies through conformity to a minimum set of ontological requirements.

In an evaluation of the *International Classification of Health Interventions* (ICHI), which is one of the reference classifications in the WHO-FIC network, in the coding of common surgical procedures, Fung et al. (2022) identified the 300 most performed surgical procedures, as represented by their display names in an electronic health record and coded with multiple KOSs. The findings considered that post-coordination (modification of existing codes by adding other codes) could be an effective and efficient means of improving coverage.

8.3.2 Making health KOSs sharable and reusable

In an effort to identify the next set of goals for the development of standards, making controlled vocabularies sharable and reusable, Cimino (1998) generated a list of the thoughts regarding additional characteristics needed according to the common themes found in the publications by medical informatics researchers. The categories include vocabulary content, concept orientation, concept permanence, non-semantic concept identifiers, polyhierarchy, formal definitions, rejection of “not elsewhere classified” terms, multiple granularities, multiple consistent views, context representation, graceful evolution, and recognized redundancy.

Researchers have been discussing the need for a common or shared ontology when handling the KOSs that have differences not only in contents but also in the fundamental structures, such as a classification (ICD) and a clinical reference terminology (SNOMED CT), to ensure seamless reuse and semantic interoperability (Rodrigues et al. 2013). ICD and SNOMED CT are the two most important and widely used KOSs to normalize the diagnostic process and health problems list in EHR (refer to Figure 7 in Section 7.1). The differences are seen in their nature of term coverage (e.g., over 100,000 concepts in SNOMED cover the same area as ICD-10 with the most used terms, according to Chavis 2013), and compositional nature, such as for hierarchical and associative relationships (Chute et al. 1996, Campbell et al. 1997). As Rodrigues et al. (2015) advised, due to fundamental differences in design and editorial policies, semantic interoperability between them requires combining two different approaches: 1) axiom-based, which states logically what is universally true using an ontology language such as OWL, and 2) rule-based, expressed as queries on the axiom-based knowledge.

The nature of multiple KOSs has made the medical terminologies and mappings across them crucial pre-requisites for achieving interoperable eHealth applications (Hussain et al. 2014). The mapping processes and focuses can vary, as categorized by Iyer and Giguere (1995), involving: exact matches, specific to general, general to specific, many to one, cyclic mapping strategies, no matches, and specific and broad class mapping. In addition to the processes, a framework was designed by a cross-country team (Hussain et al. 2014) to enable a more collaborative semantic landscape with providers and consumers of terminology mappings, built upon existing terminology mappings to (i) infer new mappings for different computable semantic interoperability (CSI) use cases, (ii) present provenance of the mappings together with the context information—an important problem for term mapping utilization, and (iii) perform mapping validation in order to show that inferred mappings can be erroneous.

Based on their prior work, a group of 13 co-authors documented 12 case studies of terminology mappings. In addition to the mappings of different ICD versions, ICD was mapped to others, such as Orpha-codes, SNOMED CT, ICHI, and ICF. Their prior work on mapping these KOSs addressing specific vocabularies, health areas, issues, and tool development has been published in *Journal of the American Medical Informatics Association*, *Journal of Biomedical Informatics*, *International Journal of Medical Informatics*, and the series *Studies in Health Technology and Informatics*. Their solid studies and experiments led to a technical document released by the WHO Classifications and Terminologies team entitled *WHO-FIC Classifications and Terminology Mapping – Principles and Best Practice* [2021]. These best practices are grounded in the case study authors’ experiences and supplement those top-down principles enumerated in works such as *ISO/TR 12300:2014 International Organization for Standardization. Health Informatics—Principles of Mapping between Terminological Systems*.

To address interoperability challenges in heterogeneous data sources, a collaborative study by Ethier et al. (2013) developed a framework based on the general information model (GIM) ontology, to unify structural models and terminologies together with relevant mapping sets. The modeling infrastructure resides entirely within a terminology server, enabling unification of structural and semantic modeling and operations within this server. Several types of models are present: 1. The GIM (general information model), 2. Models describing each data source (DSM), 3. Mapping sets between the sources and the GIM—one set per source, 4. Terminologies used to code the data elements (e.g., ICD-10 codes), and 5. Mappings between terminologies. The unified ontology-based knowledge framework has been tested in the context of the EU Framework Program 7 TRANSFoRM project, where it was used to achieve data integration in a retrospective diabetes cohort study. It is obvious that systems using the architecture can rely solely on GIM concepts, abstracting over both the structure and coding specificities of the data sources to facilitate interoperability between heterogeneous sources.

ICD version transition and code conversion is another noticeable theme. ICD’s revisions directly impact all health workflows, since any changing processes related to controlled vocabularies must be reconciled with historical patient information, which is coded using those vocabularies and stored in clinical databases. These changes observed in ICD and other controlled vocabularies include adding terms, deleting terms, changing terms, and other special situations such as obsolescence, discovering redundancy, and pre-coordination (Cimino and Clayton 1994). The effects of vocabulary mapping have been studied from a health informatics perspective, for example, for conditions on phe-

notype cohorts (Hripcsak et al. 2018). On the other hand, Nadkarni and Darer (2010) pointed out that strategies to deal with legacy ICD data must also address the issue of codes created by non-taxonomist users.

9.0 Conclusion

Since many communities, professions, and subject disciplines have developed different ways to classify things and organize knowledge, the high-level effort of promoting a unified classification system such as ICD is critical, as it provides a common language for reporting and monitoring diseases and has been used worldwide for morbidity and mortality statistics. Furthermore, by encoding the diseases and representing the concepts in a standard way, ICD greatly enhances the constancy of coding across languages, cultures, and healthcare systems around the world. The timely actions and guidelines provided by ICD during the COVID-19 pandemic have been followed by other standard health KOSs all over the world (Zeng, Hong et al. 2020).

The focus of this paper is to present ICD as an important KOS, demonstrated by its wide usage, adaption, extension, revision, translation, and the direct implementations of its updates. The ICD has been developed as a practical, rather than a purely theoretical classification (World Health Organization 2022a, Sec. 1.2.1). The generation and modification of the ICD revisions has presented significant changes, as visible from the examples presented in this paper, in the current versions ICD-10 and ICD-11. Another noteworthy practice that is somewhat different from normal classifications is its notation systems, which contain different versions for both semantic and ordinal values that could directly impact the functionality, consistency, and accuracy of the administrative data along with its interoperability with other KOSs (Gnoli 2018, Iyer 1995). The newest edition, ICD-11, represented multiple innovative approaches to serving semantic interoperability of individual data and reusability of recorded data, in order to allow the systematic recording, analysis, interpretation, and comparison of mortality and morbidity data collected in different countries or regions and at different times (World Health Organization 2022a, Sec. 1.1).

Acknowledgements

The authors would like to thank the editor-in-chief of the *ISKO Encyclopedia of Knowledge Organization*, Prof. Birger Hjørland, and the two anonymous referees for providing their valuable feedback, including detailed, knowledgeable, and fruitful suggestions for improving the original manuscript.

Endnotes

1. In the full article “Florence Nightingale and the early origins of evidence-based nursing”, Lynn McDonald (2001) indicated that “Nightingale returned from the Crimean War with a conviction that the desperate loss of life she witnessed should never occur again. She pressed for the creation of a royal commission to investigate the causes of the high mortality (eg, for every 1 soldier that died from his wounds, 7 died from disease). She worked with a team of “sanitary experts,” including William Farr, who was Britain’s leading social statistician at the time.” <http://dx.doi.org/10.1136/ebn.4.3.68>
2. Classification of diseases by anatomical site is common, and, for example, reflected in medical specialties (such as cardiologists, dermatologists, neurologists, ophthalmologists etc.) But this does not mean that classification of diseases by anatomical site is necessarily the optional classification. Berman (2005, 2), for example, wrote: “Pathologists typically refer to anatomic tumor classifications when they are more accurately referring to lists of primary tumors that are known to occur at a particular location [references here omitted]. A list of tumors occurring at a body site is not a classification because it includes tumors that are biologically, clinically, and histologically unrelated. Although often referred to as World Health Organization (WHO) “classifications”, the WHO accurately titles their organ-based lists of neoplasms as “Histologic Typings” for the different organs [references here omitted].”
3. Wolfbane Cybernetic Ltd [<http://www.wolfbane.com/>] provides a set of archives on its website, including the archive for ICD, from ICD Revision 1 to 10, based on its own digitized products. <http://www.wolfbane.com/icd/index.html>. It is used in this article for early ICD revision’s resources.
4. The World Health Organization (2022a) *ICD-11 Reference Guide* Section 2.2 defines “[t]he Tabular list is an alphanumeric listing of diseases and disease groups, inclusion and exclusion notes, and some coding rules.” The Glossary of *Diagnosis Coding: Using the ICD-10-CM*, by Centers for Medicare & Medicaid Services (CMS), defines “Tabular List of Diseases and Injuries” as: “[a] structured list of codes divided into chapters based on body system or condition.” <https://www.cms.gov/Outreach-and-Education/MLN/WBT/MLN6447308-ICD-10-CM/icd10cm/misc/glossary/index.html>.
5. For Aristotelian Framework, consult chapter 2 in: Richards, Richard A. 2016. *Biological Classification: A Philosophical Introduction*. Cambridge: Cambridge University Press.

References

- ANSI/NISO Z39.19-2005 (R2010) *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Baltimore, MD: National Information Standards Organization.
- Ben-Ezra, Menachem, Wai Kai Hou and Robin Goodwin. 2021. "Investigating the Relationship between COVID-19-Related and Ristress and ICD-11 Adjustment Disorder: Two Cross-Sectional Studies." *BJPsych Open* 7(1): 1-7. doi: 10.1192/bjo.2020.158
- Berman, Jules. 2005. "Modern Classification of Neoplasms: Reconciling Differences between Morphologic and Molecular Approaches." *BMC Cancer* 5:. doi: 10.1186/1471-2407-5-100
- Berners-Lee, Tim, Roy Fielding and Larry Masinter. 1998. "Uniform Resource Identifiers (URI): Generic Syntax." Retrieved from: <https://datatracker.ietf.org/doc/html/rfc3986>
- Bhatt, Ankeet S., Erin E. McElrath, Brian L. Claggett, Deepak L. Bhatt, Dale S. Adler, Scott D. Solomon and Muthiah Vaduganathan. 2021. "Accuracy of ICD-10 Diagnostic Codes to Identify COVID-19 Among Hospitalized Patients." *Journal of General Internal Medicine* 36: 2532-5.
- Blosnich, John R., John Cashy, Adam J. Gordon, Jillian C. Shipherd, Michael R. Kauth, George R. Brown and Michael J. Fine. 2018. "Using Clinician Text Notes in Electronic Medical Record Data to Validate Transgender-Related Diagnosis Codes." *Journal of the American Medical Informatics Association* 25: 905-8.
- Boslett, Andrew J., Alina Denham, Elaine L. Hill and Meredith CB Adams. 2019. "Unclassified Drug Overdose Deaths in the Opioid Crisis: Emerging Patterns of Inequity." *Journal of the American Medical Informatics Association* 26: 767-77.
- Bowker, Geoffrey and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- Brown, Philip J. B. and Peter Sönksen. 2000. "Evaluation of the Quality of Information Retrieval of Clinical Findings from a Computerized Patient Database Using a Semantic Terminological Model." *Journal of the American Medical Informatics Association* 7: 392-403.
- Campbell, Elizabeth A., Ellen J. Bass and Aaron J. Masino. 2020. "Temporal Condition Pattern Mining in Large, Sparse Electronic Health Record Data: A Case Study in Characterizing Pediatric Asthma." *Journal of the American Medical Informatics Association* 27: 558-66.
- Campbell, James R., Paul Carpenter, Charles Sneiderman, Simon Cohn, Christopher G. Chute and Judith Warren. 1997. "Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions, and Clarity." *Journal of the American Medical Informatics Association* 4: 238-51.
- Cervone, H. Frank. 2018. "An Overview of ICD Medical Classification for the Information Professional." *Digital Library Perspectives* 34: 2-8.
- Chan, Lois Mai and Athena Salaba. 2016. *Cataloging and Classification: An Introduction*. 4th ed. Lanham, MD: Rowman & Littlefield.
- Chute, Christopher G., Simon P. Cohn, Keith E. Campbell, Diane E. Oliver, James R. Campbell and Computer-Based Patient Record Institute's Work Group on Codes & Structures. 1996. "The Content Coverage of Clinical Classifications." *Journal of the American Medical Informatics Association* 3: 224-33.
- Cimino, James J. 1998. "Desiderata for Controlled Medical Vocabularies in the Twenty-First Century." *Methods of information in medicine* 37: 394-403.
- Cimino, James J. and Paul D. Clayton. 1994. "Coping with Changing Controlled Vocabularies." In *Proceedings of the Annual Symposium on Computer Application in Medical Care, 1994*: 135-9.
- CMS Medicare Learning Network. 2009. "General Equivalency Mappings - ICD-9-CM to and from ICD-10-CM and ICD-10-PCS." https://www.cms.gov/medicare/coding/icd10/downloads/icd-10_gem_fact_sheet.pdf
- Comaromi, John Phillip. 1976. *The Eighteen Editions of the Dewey Decimal Classification*. Albany, NY: Forest Press.
- Denaxas, Spiros, Arturo Gonzalez-Izquierdo, Kenan Direk, Natalie K. Fitzpatrick, Ghazaleh Fatemifar, Amitava Banerjee, Richard JB Dobson, Laurence J Howe, Valerie Kuan, R Tom Lumbers, Laura Pasea, Riyaz S Patel, Anoop D Shah, Aroon D Hingorani, Cathie Sudlow and Harry Hemingway. 2019. "UK Phenomics Platform for Developing and Validating Electronic Health Record Phenotypes: CALIBER." *Journal of the American Medical Informatics Association* 26: 1545-59.
- Dewey Services. [2017]. "A legacy of helping libraries." OCLC. Retrieved from: <https://www.oclc.org/en/dewey/resources/timeline.html>
- DiChiara, Jacqueline. 2015. "ICD-1 to ICD-11 Timeline Highlights Healthcare's Evolution." In: *Xtelligent Healthcare Media Policy & Regulation News*, August 11, 2015. Retrieved from: <https://revcycleintelligence.com/news/icd-1-to-icd-11-timeline-highlights-healthcares-evolution>
- Drösler, Saskia E., Stefanie Weber and Christopher G. Chute. 2021. "ICD-11 Extension Codes Support Detailed Clinical Abstraction and Comprehensive Classification." *BMC Medical Informatics and Decision Making* 21, no. 6: 1-9.
- Eisma, Maarten C., Aerjen Tamminga, Geert E. Smid and Paul A. Boelen. 2021. "Acute Grief after Deaths due to COVID-19, Natural Causes and Unnatural Causes: An

- Empirical Comparison.” *Journal of Affective Disorders* 278: 54-6.
- Encyclopedia.com. 2019. “Disease, Concepts and Classification of.” *Encyclopedia of Population*. Retrieved from: <https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/disease-concepts-and-classification>
- Ethier, Jean-François, Olivier Dameron, Vasa Curcin, Mark M. McGilchrist, Robert A. Verheij, Theodoros N. Arvanitis, Adel Taweel, Brendan C. Delaney and Anita Burgun. 2013. “A Unified Structural/Terminological Interoperability Framework based on LexEVS: Application to TRANSFoRm.” *Journal of the American Medical Informatics Association* 20: 986-94.
- Eyler, John M. 1979. *Victorian Social Medicine: The Ideas and Methods of William Farr*. Baltimore: Johns Hopkins University Press.
- Fortune, Nicola, Richard Madden and Ann-Helene Almborg. 2018. “Use of a New International Classification of Health Interventions for Capturing Information on Health Interventions Relevant to People with Disabilities.” *International Journal of Environmental Research and Public Health* 15: 145. doi:10.3390/ijerph15010145
- Fung, Kin Wah, Julia Xu, Filip Ameye, Lisa Burelle and Janice MacNeil. 2022. “Evaluation of the International Classification of Health Interventions (ICHI) in the Coding of Common Surgical Procedures.” *Journal of the American Medical Informatics Association* 29: 43-51.
- Gao, Shang, Michael T. Young, John X. Qiu, Hong-Jun Yoon, James B. Christian, Paul A. Fearn, Georgia D. Tourassi and Arvind Ramanathan. 2018. “Hierarchical Attention Networks for Information Extraction from Cancer Pathology Reports.” *Journal of the American Medical Informatics Association* 25: 321-30.
- Gerbier-Colomban, Solweig, Véronique Potinet-Pagliaroli and Marie-Hélène Metzger. 2014. “Can Epidemic Detection Systems at the Hospital Level Complement Regional Surveillance Networks: Case Study with the Influenza Epidemic?” *BMC Infectious Diseases* 14: 1-10.
- Gnoli, Claudio. 2018. “Notation.” *Knowledge Organization* 45: 667-84. Also available in *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjørland and Claudio Gnoli. <http://www.isko.org/cyclo/notation>
- Grove, John and Robert Jakob. 2018. “ICD 11th Revision.” Member State Information Session, Geneva, 14 May 2018, World Health Organization. <https://dxrevisionwatch.files.wordpress.com/2018/05/icd11.pdf>
- Harrison, James E., Stefanie Weber, Robert Jakob and Christopher G. Chute. 2021. “ICD-11: An International Classification of Diseases for the Twenty-First Century.” *BMC Medical Informatics and Decision Making* 21, no. 6: 1-10.
- Hjørland, Birger. 2016. “Knowledge Organization.” *Knowledge Organization* 43: 475-84. Also available in *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjørland and Claudio Gnoli. http://www.isko.org/cyclo/knowledge_organization
- Hjørland, Birger. 2017. “Classification.” *Knowledge Organization* 44: 97-128. Also available in *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjørland and Claudio Gnoli. <http://www.isko.org/cyclo/classification>
- Hjørland, Birger. 2018. “Indexing: Concepts and Theory.” *Knowledge Organization* 45: 609-39. Also available in *ISKO Encyclopedia of Knowledge Organization*, edited by Birger Hjørland and Claudio Gnoli. <http://www.isko.org/cyclo/indexing>
- HL7 International. 2007-2022. “HL7 Version 2 Product Suite.” Retrieved from: https://www.hl7.org/implementation/standards/product_brief.cfm?product_id=185.
- Hohl, Corinne M., Andrei Karpov, Lisa Reddekopp and Jürgen Stausberg. 2014. “ICD-10 Codes Used to Identify Adverse Drug Events in Administrative Data: A Systematic Review.” *Journal of the American Medical Informatics Association* 21: 547-57.
- Hohnloser, J. H. and H. Soltanian. 1994. “Positive Efficiency Findings Using Computer Assisted ICD-Encoding: 3.5 Years of Experience with the Computerized Patient Record System PADS (Patient Archiving & Documentation System).” In *Proceedings of the Annual Symposium on Computer Application in Medical Care, 1994*: 965-6.
- Horsky, Jan, Elizabeth A. Drucker and Harley Z. Ramelson. 2017. “Accuracy and Completeness of Clinical Coding using ICD-10 for Ambulatory Visits.” In *AMIA Annual Symposium Proceedings, 2017*. American Medical Informatics Association, 912-20.
- Hripcsak, George, Matthew E. Levine, Ning Shang and Patrick B. Ryan. 2018. “Effect of Vocabulary Mapping for Conditions on Phenotype Cohorts.” *Journal of the American Medical Informatics Association* 25:1618-25.
- Hunter, Eric J. 2009. *Classification Made Simple: An Introduction to Knowledge Organisation and Information Retrieval*. 3rd ed. Farnham: Ashgate Publishing.
- Hussain, Sajjad, Hong Sun, Anil Sinaci, Gokce Banu Laleci Erturkmen, Charles Mead, Alasdair JG Gray, Deborah L. McGuinness, Eric Prud’Hommeaux, Christel Daniel and Kerstin Forsberg. 2014. “A Framework for Evaluating and Utilizing Medical Terminology Mappings.” *Studies in Health Technology and Informatics* 205: 594-8.
- International Committee on Taxonomy of Viruses Coronavirus Study Group. 2020. “The Species *Severe Acute Respiratory Syndrome-related Coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2.” *Nature Mi-*

- crobiology* 5: 536-44. <https://doi.org/10.1038/s41564-020-0695-z>
- ISO 17115:2020 *Health Informatics—Representation of Categorical Structures of Terminology (CatStructure)*. ISO/TC 215 Health Informatics.
- ISO 1828:2012 *Health informatics—Categorical Structure for Terminological Systems of Surgical Procedures (ISO 1828:2012)*. ISO/TC 215 Health Informatics.
- ISO/TR 12300:2014 *International Organization for Standardization. Health Informatics—Principles of Mapping between Terminological Systems*. ISO/TC 215 - Health Informatics ISO/TC 215/WG 3 - Semantic Content.
- Iyer, Hemalata. 1995. *Classificatory Structures: Concepts, Relations and Representation*. Frankfurt am Main: Indeks Verlag.
- Iyer, Hemalata and Mark Giguere. 1995. "Towards Designing an Expert System to Map Mathematics Classificatory Structures." *Knowledge Organization* 22: 141-7.
- Jetté, Nathalie, Hude Quan, Brenda Hemmelgarn, Saskia Drösler, Christina Maass, Lori Moskal, Wansa Paoim, Vijaya Sundararajan, Song Gao, Robert Jakob, Bedihran Üstün, William A. Ghali and the IMECCHI Investigators. 2010. "The Development, Evolution, and Modifications of ICD-10: Challenges to the International Comparability of Morbidity Data." *Medical Care* 48: 1105-10.
- Keck, James W., John T. Redd, James E. Cheek, Larry J. Layne, Amy V. Groom, Sassa Kitka, Michael G. Bruce, Anil Suryaprasad, Nancy L. Amerson, Theresa Cullen, Ralph T. Bryan and Thomas W. Hennessy. 2014. "Influenza Surveillance Using Electronic Health Records in the American Indian and Alaska Native Population." *Journal of the American Medical Informatics Association* 21: 132-8.
- Killikelly, C., G. E. Smid, B. Wagner and P. A. Boelen. 2021. "Responding to the New International Classification of Diseases-11 Prolonged Grief Disorder During the COVID-19 Pandemic: A New Bereavement Network and Three-Tiered Model of Care." *Public Health* 191: 85-90.
- Knibbs, George H. 1929. "The International Classification of Disease and Causes of Death and its Revision." *Medical Journal of Australia* 1: 2-12.
- Kronk, Clair A. and Judith W. Dexheimer. 2020. "Development of the Gender, Sex, and Sexual Orientation Ontology: Evaluation and Workflow." *Journal of the American Medical Informatics Association* 27: 1110-5.
- Kullo, Iftikhar J., Jin Fan, Jyotishman Pathak, Guergana K. Savova, Zeenat Ali and Christopher G. Chute. 2010. "Leveraging Informatics for Genetic Studies: Use of the Electronic Medical Record to Enable a Genome-Wide Association Study of Peripheral Arterial Disease." *Journal of the American Medical Informatics Association* 17: 568-74.
- Lewes, F. M. M. 1988. "Dr Marc d'Espine's Statistical Nomenclology." *Medical History* 32: 301-3.
- Lilienfeld, David E. 2007. "Celebration: William Farr (1807-1883) - An Appreciation on the 200th Anniversary of His Birth." *International Journal of Epidemiology* 36: 985-7.
- Livingstone-Banks, Jonathan. 2018. "The Case for a Meta-Nosological Investigation of Pragmatic Disease Definition and Classification." *Journal of Evaluation in Clinical Practice* 24: 1013-8.
- Loukides, Grigorios, Joshua C. Denny and Bradley Malin. 2010. "The Disclosure of Diagnosis Codes can Breach Research Participants' Privacy." *Journal of the American Medical Informatics Association* 17: 322-7.
- Madden, Richard, Catherine Sykes and T. Bedirhan Ustun. [2007]. "World Health Organization Family of International Classifications: Definition, Scope and Purpose." Geneva: World Health Organization. <https://www.who.int/classifications/en/FamilyDocument2007.pdf>
- Mahajan, Reena, Anne C. Moorman, Stephen J. Liu, Loralee Rupp, R. Monina Klevens and Chronic Hepatitis Cohort Study (CHeCS) Investigators. 2013. "Use of the International Classification of Diseases, 9th Revision, Coding in Identifying Chronic Hepatitis B Virus Infection in Health System Data: Implications for National Surveillance." *Journal of the American Medical Informatics Association* 20: 441-5.
- Mamykina, Lena, David K. Vawdrey, Peter D. Stetson, Kai Zheng and George Hripcsak. 2012. "Clinical Documentation: Composition or Synthesis?" *Journal of the American Medical Informatics Association* 19: 1025-31.
- Manchaiah, Vinaya, De Wet Swanepoel and Nicola Fortune. 2020. "The International Classification of Health Interventions (ICHI) - A New Tool for Describing and Reporting Interventions in Audiology." *International Journal of Audiology* 59: 403-5.
- McDonald, Lynn. 2001. "Florence Nightingale and the Early Origins of Evidence-Based Nursing." *Evidence-Based Nursing* 4, no. 3: 68-9.
- Mezzich, Juan. 2002. "International Surveys on the Use of ICD-10 and Related Diagnostic Systems." *Psychopathology* 35, nos. 2-3: 72-5.
- Moore, Karen, James Black, Stacey Rowe and Lucinda Franklin. 2011. "Syndromic Surveillance for Influenza in Two Hospital Emergency Departments. Relationships between ICD-10 Codes and Notified Cases, Before and During a Pandemic." *BMC Public Health* 11: 1-7.
- Moriyama, Iwao M, Ruth M. Loy, Alastair H. T. Robb-Smith, Harry M. Rosenberg and Donna L. Hoyert. 2011. *History of the Statistical Classification of Diseases and Causes of Death*. Hyattsville, MD: National Center for Health Statistics, Centers for Disease Control and

- Prevention (CDC). https://www.cdc.gov/nchs/data/misc/classification_diseases2011.pdf
- Nadkarni, Prakash M. and Jonathan A. Darer. 2010. "Migrating Existing Clinical Content from ICD-9 to SNOMED." *Journal of the American Medical Informatics Association* 17: 602-7.
- National Center for Biomedical Ontology. 2005-. "About NCBO." Retrieved from: <https://ncbo.bioontology.org/about-ncbo>
- Patel, Vimla L., Andre W. Kushniruk, Seungmi Yang and Jean-Francois Yale. 2000. "Impact of a Computer-based Patient Record System on Data Collection, Knowledge Organization, and Reasoning." *Journal of the American Medical Informatics Association* 7: 569-85.
- Perotte, Adler, Rimma Pivovarov, Karthik Natarajan, Nicole Weiskopf, Frank Wood and Noémie Elhadad. 2014. "Diagnosis Code Assignment: Models and Evaluation Metrics." *Journal of the American Medical Informatics Association* 21: 231-7.
- Phull, Jaspreet. 2012. *ICD 10 in Psychiatry: A Learning Guide*. London: CRC Press.
- Proctor, B. Deborah and Alexandra Patricia Young-Adams. 2014. "Basics of Diagnostic Coding." Chapter 18 in *Kinn's the Medical Assistant: An Applied Learning Approach*. 12th ed. St. Louis, MO: Elsevier.
- Ranganathan, Shiyali Ramamrita. 1967. *Prolegomena to Library Classification*. 3rd ed. Bombay: Asia Publishing House.
- Robb-Smith, A. H. T. 1969. "A History of the College's Nomenclature of Diseases: Its Preparation." *Journal of the Royal College of Physicians of London* 3: 341-58.
- Rodrigues, Jean M., Ann Casey, Cédric Bousquet, Anand Kumar, Pierre Lewalle and Béatrice Trombert Paviot. 2011. "Comparing Existing National and International Classification Systems of Surgical Procedures with the CEN/ISO 1828 Ontology Framework Standard." In *User Centred Networked Health Care. Proceedings of MIE 2011*. IOS Press, 844-8.
- Rodrigues, Jean-Marie, David Robinson, Vincenzo Della Mea, James Campbell, Alan Rector, Stefan Schulz, Hazel Brear et al. 2015. "Semantic Alignment between ICD-11 and SNOMED CT." In *MEDINFO 2015: eHealth-enabled Health*, edited by I.N. Sarkar et al. IMIA and IOS Press, 790-4.
- Rodrigues, Jean-Marie, Stefan Schulz, Alan Rector, Kent Spackman, Bedirhan Üstün, Christopher G. Chute, Vincenzo Della Mea, Jane Millar and Kristina Brand Persson. 2013. "Sharing Ontology between ICD 11 and SNOMED CT will Enable Seamless Re-use and Semantic Interoperability." In *MEDINFO 2013*, edited by C.U. Lehmann et al. IMIA and IOS Press, 343-6.
- Rosenbloom, S. Trent, William W. Stead, Joshua C. Denny, Dario Giuse, Nancy M. Lorenzi, Steven H. Brown and Kevin B. Johnson. 2010. "Generating Clinical Notes for Electronic Health Record Systems." *Applied Clinical Informatics* 1: 232-43.
- Salcido, Richard. 2015. "The ABCs of ICD." *Advances in Skin & Wound Care* 28, no. 2: 53.
- Schiff, Gordon D., David W. Bates, P. Hartzband, J. Groopman and G. D. Schiff. 2010. "Can Electronic Clinical Documentation Help Prevent Diagnostic Errors?" *New England Journal of Medicine* 362: 1066-9.
- Severinsen, Morten. 2001. "Principles Behind Definitions of Diseases – a Criticism of the Principle of Disease Mechanism and the Development of a Pragmatic Alternative." *Theoretical Medicine and Bioethics* 22: 319-36.
- Smith, Edward E. and Douglas L. Medin. 1981. *Categories and Concepts*. Cambridge, MA: Harvard University Press.
- Terron Cuadrado, Marta. 2019. "National Modifications of WHO ICD." In *eHealth DSI Semantic Community Home – ASSETS*. Retrieved from: <https://ec.europa.eu/cefdigital/wiki/display/EHSEMANTIC/National+modifications+of+WHO+ICD>. (Last modified on Feb 01, 2019.)
- Thompson, Edward T. and Adaline C. Hayden (Eds.). 1961. *Standard Nomenclature of Diseases Operations*. New York: McGraw-Hill.
- Tian, Terrence Y., Ianita Zlateva and Daren R. Anderson. 2013. "Using Electronic Health Records Data to Identify Patients with Chronic Pain in a Primary Care Setting." *Journal of the American Medical Informatics Association* 20, no. e2: e275-80.
- Tsui, Fu-Chiang, Michael M. Wagner, Virginia Dato and Chung-Chou Ho Chang. 2002. "Value of ICD-9-Coded Chief Complaints for Detection of Epidemics." *Journal of the American Medical Informatics Association* 9, no. Supplement_6: S41-7.
- U.S. Bureau of the Census. 1911. *Manual of the International List of Causes of Death: Based on the Second Decennial Revision by the International Commission, Paris, July 1 to 3, 1909*. Washington, DC: US Government Printing Office.
- U.S. Department of Health, Education and Welfare. 1959. *International Classification of Diseases: Adapted for Indexing of Hospital Records and Operation Classification*. US Government Printing Office.
- U.S. Department of Health, Education and Welfare. 1968. *Eighth Revision International Classification of Diseases, Adapted for Use in the United States: Alphabetical index*. Vol. 2. US Government Printing Office.
- U.S. National Center for Health Statistics. 2021. "ICD-9-CM." Retrieved from: <https://www.cdc.gov/nchs/icd/icd9cm.htm>
- U.S. National Library of Medicine. 2016-. "Source Vocabularies." In *UMLS Basics - Metathesaurus*. Retrieved

- 2022-06-26 from: https://www.nlm.nih.gov/research/umls/new_users/online_learning/Meta_002.html
- U.S. National Library of Medicine. 2022. "Statistics - 2022AA Release." In UMLS Metathesaurus Release Documentation. Retrieved 2022-06-26 from: https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/statistics.html
- Wang, Liqin, Maxim Topaz, Joseph M. Plasek and Li Zhou. 2017. "Content and Trends in Medical Informatics Publications Over the Past Two Decades." *MedInfo* 245: 968-72.
- Wei, Wei-Qi, Pedro L. Teixeira, Huan Mo, Robert M. Cronin, Jeremy L. Warner and Joshua C. Denny. 2016. "Combining Billing Codes, Clinical Notes, and Medications from Electronic Health Records Provides Superior Phenotyping Performance." *Journal of the American Medical Informatics Association* 23, no. e1: e20-7.
- WHO-FIC Classifications and Terminology Mapping - Principles and Best Practice. [2021]. https://cdn.who.int/media/docs/default-source/classification/who-fic-network/whofic_terminology_mapping_guide.pdf
- Wiley, Laura K., Anushi Shah, Hua Xu and William S. Bush. 2013. "ICD-9 Tobacco Use Codes are Effective Identifiers of Smoking Status." *Journal of the American Medical Informatics Association* 20: 652-8.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*, translated by G.E.M. Anscombe and edited by G.E.M. Anscombe and R. Rhees. Oxford: Blackwell.
- Wolfbane Cybernetic. 2001-. International Classification of Diseases. Updated 7 April 2007. Accessed 2022-06-26 from: <http://www.wolfbane.com/icd/index.html>
- World Health Organization. 1948. *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death: Sixth Revision of the International Lists of Diseases and Causes of Death, Adopted 1948*. World Health Organization. Available from: <https://apps.who.int/iris/handle/10665/42893>
- World Health Organization. 1978. *International Classification of Procedures in Medicine*. Geneva: World Health Organization. Available from: <https://apps.who.int/iris/handle/10665/40527>
- World Health Organization. [1996-]. "List of Official ICD-10 Updates." Available from: <https://www.who.int/standards/classifications/classification-of-diseases/list-of-official-icd-10-updates>
- World Health Organization. [2004]. *History of the Development of the ICD*. <https://cdn.who.int/media/docs/default-source/classification/icd/historyoficd.pdf>
- World Health Organization. 2007-. *International Classification of Health Interventions (ICHI)*. Geneva: World Health Organization. Retrieved from: <https://www.who.int/standards/classifications/international-classification-of-health-interventions>
- World Health Organization. 2015. "WHO Best Practices for Naming of New Human Infectious Diseases." Retrieved from: https://www.who.int/topics/infectious_diseases/naming-new-diseases/en/
- World Health Organization. 2016. *International Statistical Classification of Diseases and Related Health Problems. - 10th revision, Fifth edition*. Geneva: World Health Organization. Available from: <https://apps.who.int/iris/handle/10665/246208>
- World Health Organization. 2018. "WHO Releases New International Classification of Diseases (ICD 11)." News release. Geneva: World Health Organization. Retrieved from: [https://www.who.int/news/item/18-06-2018-who-releases-new-international-classification-of-diseases-\(icd-11\)](https://www.who.int/news/item/18-06-2018-who-releases-new-international-classification-of-diseases-(icd-11))
- World Health Organization. 2019a. "World Health Assembly Update, 25 May 2019." In WHO News, 25 May 2019. Retrieved from: <https://www.who.int/newsroom/detail/25-05-2019-world-health-assembly-update>
- World Health Organization. 2019b. "ICD-10 Version:2019." <https://icd.who.int/browse10/2019/en>
- World Health Organization. 2019c. *ICD-11 Implementation or Transition Guide*. https://icd.who.int/docs/ICD-11%20Implementation%20or%20Transition%20Guide_v105.pdf
- World Health Organization. 2020a. "ICD-10-CM Official Coding Guidelines - Supplement Coding Encounters Related to COVID-19 Coronavirus Outbreak. Effective: February 20, 2020." <https://www.cdc.gov/nchs/data/icd/ICD-10-CM-Official-Coding-Guidance-Interim-Advice-coronavirus-feb-20-2020.pdf>
- World Health Organization. 2020b. "COVID-19 Coding in ICD-10." 25 March 2020. <https://www.who.int/classifications/icd/COVID-19-coding-icd10.pdf>
- World Health Organization. 2021-. "ICD-11 Fact Sheet." <https://www.who.int/publications/m/item/icd-11-fact-sheet>
- World Health Organization. 2021a. "History of ICD." In "International Statistical Classification of Diseases and Related Health Problems (ICD)" homepage, retrieved from: <https://www.who.int/standards/classifications/classification-of-diseases>
- World Health Organization. 2021b. "WHO Announces Simple, Easy-To-Say Labels for SARS-CoV-2 Variants of Interest and Concern." In WHO News, 31 May 2021. Retrieved from: <https://www.who.int/news/item/31-05-2021-who-announces-simple-easy-to-say-labels-for-sars-cov-2-variants-of-interest-and-concern>
- World Health Organization. 2021c. "It All Starts With A Code ..." <https://cdn.who.int/media/images/default-source/classifications/icd/covid-flowchart-final-tmb-1920v.png>. In *Emergency Use ICD Codes for COVID-19 Disease Outbreak*. Retrieved from: <https://www.who.int>

int/standards/classifications/classification-of-diseases/emergency-use-icd-codes-for-covid-19-disease-outbreak

World Health Organization. 2021d. "World Health Organization Family of International Classifications 2021." <https://cdn.who.int/media/docs/default-source/classification/who-fic-network/who-fic-family-paper.pdf>

World Health Organization. 2022a. *ICD-11 Reference Guide*. https://icdcdn.who.int/icd11referenceguide/en/ICD-11_Reference_guide_2022_en.pdf (An HTML version is available in <https://icd.who.int/icd11refguide/en/index.html>)

World Health Organization. 2022b. "ICD-11 2022 Release." In *WHO News*, 11 February 2022. <https://www.who.int/news/item/11-02-2022-icd-11-2022-release>

Yan, Chenwei, Xiangling Fu, Xien Liu, Yuanqiu Zhang, Yue Gao, Ji Wu, and Qiang Li. 2022. "A Survey of Automated ICD Coding: Development, Challenges, and Applications." *Intelligent Medicine 2*, doi: <https://doi.org/10.1016/j.imed.2022.03.003>

Zeng, Marcia Lei, Yi Hong, Julaine Clunis, Shaoyi He and L. P. Coladangelo. 2020. "Implications of Knowledge Organization Systems for Health Information Exchange and Communication During the COVID-19 Pandemic." *Data and Information Management 4*: 148-70.

Appendix:

List of health KOS standards presented in this article

KOS	Full title	Description	Edition / Year	Reference of Full Version; Online Browser
CPT	Current Procedural Terminology	A uniform language for coding medical services and procedures to streamline reporting, increase accuracy and efficiency, as well as supporting administrative management such as claims processing and developing guidelines for medical care review.	2022	American Medical Association. 2021. <i>CPT 2022</i> . Washington, D.C.: American Medical Association. ISBN 9781640160880 / ISBN 1640160884. https://www.ama-assn.org/amaone/cpt-current-procedural-terminology
DO	Disease Ontology	A standardized ontology for human disease with the purpose of providing the biomedical community with consistent, reusable, and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts.	2022	Institute of Genome Sciences. 2022. <i>Disease Ontology</i> . Baltimore, MD: University of Maryland School of Medicine. https://disease-ontology.org
ICD-9	International classification of diseases, 9th revision	The Ninth Revision of International Classification of Diseases, the international classification standard to define the universe of diseases, disorders, injuries, and other related health conditions.	9th revision / 1977	World Health Organization. 1977. <i>International classification of diseases</i> . Geneva: World Health Organization.
ICD-9-CM	International classification of diseases, 9th revision, clinical modification	The Clinical Modification of ICD-9, an official classification system for assigning codes to diagnoses and procedures associated with hospital utilization in the United States.	9th revision, Clinical modification / 2012	ICD-9-CM Coordination and Maintenance Committee (U.S.); National Center for Health Statistics (U.S.); Centers for Medicare & Medicaid Services (U.S.). 2012. <i>ICD-9-CM: volumes 1, 2 & 3. (2012)</i> . Jackson, WY: Teton Data Systems. https://icd.codes/icd9cm
ICD-10	International statistical classification of diseases and related health problems, 10th revision	The Tenth Revision of International Classification of Diseases, the international classification standard to define the universe of diseases, disorders, injuries, and other related health conditions.	10th revision, 5th edition / 2016	World Health Organization. 2016. <i>International Statistical Classification of Diseases and Related Health Problems</i> . Geneva: World Health Organization. https://icd.who.int/browse10

KOS	Full title	Description	Edition / Year	Reference of Full Version; Online Browser
ICD-10-CM	International Statistical Classification of Diseases and Related Health Problems, 10th revision, Clinical Modification	The Clinical Modification of ICD-10, an official classification system for assigning diagnosis codes in the United States of America.	10th revision, Clinical modification / 2022	U.S. Centers for Medicare and Medicaid Services, National Center for Health Statistics. 2022. <i>International Statistical Classification of Diseases and Related Health Problems, 10th revision, Clinical Modification</i> . Baltimore, MD: Centers for Medicare and Medicaid Services; Hyattsville, MD: National Center for Health Statistics. https://icd10cmtool.cdc.gov
ICD-10-PCS	International classification of diseases, 10th revision, Procedure coding system	The Procedure coding system of ICD-10, an official classification system for inpatient procedure coding.	10th revision, Procedure coding system / 2022	Practice Management Information Corporation. 2022. <i>International classification of diseases, 10th revision, procedural coding system</i> . Los Angeles, CA: Practice Management Information Corporation (PMIC) ISBN 9781570661303 / ISBN 1570661308
ICD-11	International Classification of Diseases for Mortality and Morbidity Statistics, 11th revision	The Eleventh Revision of International Classification of Diseases, the international standard for systematic recording, reporting, analysis, interpretation and comparison of mortality and morbidity data.	11th revision / 2022	World Health Organization. 2022. <i>ICD-11 International Classification of Diseases for Mortality and Morbidity Statistics, Eleventh Revision</i> . Geneva: World Health Organization. https://icd.who.int/browse11
ICD-O-3	International Classification of Diseases for Oncology, Third Edition	An international classification standard of neoplasms, used principally in tumor or cancer registries for coding the site (topography) and the histology (morphology) of neoplasms.	3rd Edition / 2014	April G., World Health Organization. 2014. <i>International Classification of Diseases for Oncology</i> . Geneva: World Health Organization. ISBN 9789241548496
ICF	International Classification of Functioning, Disability and Health	An international classification standard of health and health-related domains in functioning and disability for measuring health and disability at both individual and population levels.	2008	World Health Organization. 2008. <i>International Classification of Functioning, Disability and Health</i> . Geneva: World Health Organization. ISBN 9789241547413 9241547413 https://icd.who.int/dev11/l-icf/en
ICHI	International Classification of Health Interventions	An international classification standard of health interventions as a common tool for reporting and analyzing health interventions for clinical and statistical purposes.	Beta-3 / 2007-	World Health Organization. 2007-. <i>International Classification of Health Interventions (ICHI)</i> . Geneva: World Health Organization. https://mitel.dimi.uniud.it/ichi
LOINC	Logical Observation Identifiers Names and Codes	An international standard for identifying health measurements, observations, and documents which is a common language for better understanding the clinical measurements and laboratory results.	V2.72 / 2022	Regenstrief Institute. 2022. "LOINC Table Files." In: <i>LOINC Table, Reports, and Users' Guide</i> . Indianapolis, IN: Regenstrief Institute. https://loinc.org/downloads/loinc-table/

KOS	Full title	Description	Edition / Year	Reference of Full Version; Online Browser
MeSH	Medical Subject Headings	A controlled and hierarchically organized vocabulary used for indexing, cataloging, and searching of biomedical and health-related information.	2022	National Library of Medicine (U.S.), National Institutes of Health (U.S.). 2022. <i>Medical Subject Headings</i> . Bethesda, MD: U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health, National Library of Medicine. https://meshb.nlm.nih.gov
RxNorm	RxNorm	A standardized nomenclature for clinical drugs, supporting semantic interoperability between drug terminologies and pharmacy management systems.	2022	National Library of Medicine (U.S.). 2022. <i>RxNorm</i> . Bethesda, MD: U.S. National Institutes of Health, National Library of Medicine. https://mor.nlm.nih.gov/RxNav
SNOMED CT	Systematized Nomenclature of Medicine—Clinical Terms	A global common language for <i>clinical terms</i> , the most comprehensive clinical terminology in use around the world as a multilingual resource with comprehensive, scientifically validated clinical content	2022	SNOMED International. 2022. <i>Systematized Nomenclature of Medicine—Clinical Terms</i> . London: SNOMED International https://browser.ihtsdotools.org